

STIC-Biotech/ChemLib

169778

From: Swope, Sheridan  
Sent: Thursday, October 27, 2005 3:29 PM  
To: STIC-Biotech/ChemLib  
Subject: 10/649,273

For 10/649,273, please search and interference search:

SID 2: regular search against the NT and AA data bases

SID 2: oligo search (274 residues) against the NT and AA data bases

SID 2: residues 148-414, regular search against the NT and AA data bases

SID 2: residues 176-414, regular search against the NT and AA data bases

Thanks!!

Sheridan Swope, Ph.D.  
Patent Examiner, AU 1656  
Recombinant Enzymes  
571-272-0943 (voice)  
E02B71 Remsen Bld (Office)  
E03C70 Remsen Bld (Mailbox)

RECEIVED  
OCT 27 2005  
STIC-BIOTECH/CHEN LIB  
(STIC)

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date completed: \_\_\_\_\_  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# \_\_\_\_\_ AA# \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

## Swope, Sheridan

---

**From:** Swope, Sheridan  
**Sent:** Wednesday, November 02, 2005 9:33 AM  
**To:** Arnold, Deirdre  
**Subject:** RE: 10/649,273

A, match of any contiguous 274 residues of sid 2

-----Original Message-----

**From:** Arnold, Deirdre  
**Sent:** Wednesday, November 02, 2005 9:10 AM  
**To:** Swope, Sheridan  
**Subject:** RE: 10/649,273

SPL FROM

Do you mean that you are looking for (A) sequences that match 274 contiguous residues of SID 2 or (B) sequences that are at least 274 residues in length and match *any size* portion of SID 2?

*Deirdre I. Arnold*

**United States Patent and Trademark Office**  
Technical Information Specialist  
STIC Biotech/Chem Library  
571-272-2532  
Remsen 1 A64

-----Original Message-----

**From:** Swope, Sheridan  
**Sent:** Wednesday, November 02, 2005 9:03 AM  
**To:** Arnold, Deirdre  
**Subject:** FW: 10/649,273

Please communicate via email, so I have a record.

"(274 residues) " means at least 274 residues.  
If that's not clear, let me know.

-----Original Message-----

**From:** Swope, Sheridan  
**Sent:** Thursday, October 27, 2005 3:29 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** 10/649,273

For 10/649,273, please search and interference search:

SID 2: regular search against the NT and AA data bases

SID 2: oligo search (274 residues) against the NT and AA data bases

SID 2: residues 148-414, regular search against the NT and AA data bases

SID 2: residues 176-414, regular search against the NT and AA data bases





# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 169776

TO: Sheridan Swope  
Location: REM/2B71/3C70  
Art Unit: 1656  
Monday, November 14, 2005

Case Serial Number: 10/649273

From: Deirdre Arnold  
Location: Biotech-Chem Library  
REM 1A64  
Phone: 571-272-2532

Deirdre.Arnold@uspto.gov

### Search Notes

Please review these results and contact me ASAP if you need to alter the search parameters.

*Please feel free to contact me if you have any questions or would like to amend the search.*

Thank you for using STIC services.

Regards,  
Deirdre Arnold



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 16:43:58 ; Search time 109.35 Seconds  
(without alignment)  
1584.102 Million cell updates/sec

Title: US-10-649-273-2  
Perfect score: 1 MLITKTAGVFPKSKRKRV.....DISKEVGASIKVQLKKEI 414  
Sequence: 1

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867879 seqs, 418409474 residues  
Total number of hits satisfying chosen parameters: 1867879

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	414	US-10-067-443-2	Sequence 2, Appli
2	2125	100.0	414	US-10-649-273-2	Sequence 2, Appli
3	2125	100.0	414	US-10-651-722-2	Sequence 2, Appli
4	2125	100.0	414	US-10-480-988-8	Sequence 8, Appli
5	2090.5	98.4	439	US-10-067-443-19	Sequence 19, Appli
6	2090.5	98.4	439	US-10-649-273-19	Sequence 19, Appli
7	2090.5	98.4	439	US-10-651-722-19	Sequence 19, Appli
8	2088	98.3	414	US-10-612-140-2	Sequence 5, Appli
9	1845	86.8	364	US-10-094-749-2039	Sequence 2039, Ap
10	1385	65.2	267	US-10-067-443-22	Sequence 22, Appli
11	1385	65.2	267	US-10-649-273-22	Sequence 22, Appli

12	1385	65.2	267	15	US-10-651-722-22	Sequence 22, Appli
13	714.5	33.6	409	20	US-11-097-143-34191	Sequence 34191, A
14	681.5	32.1	445	15	US-10-424-599-209259	Sequence 209259, A
15	634	29.8	463	14	US-10-067-443-3	Sequence 3, Appli
16	634	29.8	463	15	US-10-649-273-3	Sequence 3, Appli
17	634	29.8	463	15	US-10-651-722-3	Sequence 3, Appli
18	578	27.2	444	16	US-10-437-961-113732	Sequence 113732, A
19	556.5	26.2	333	14	US-10-012-140-25	Sequence 25, Appli
20	549	25.8	179	14	US-10-067-443-25	Sequence 25, Appli
21	549	25.8	179	15	US-10-649-273-25	Sequence 25, Appli
22	549	25.8	179	15	US-10-651-722-25	Sequence 25, Appli
23	524	24.7	382	15	US-10-282-122A-50858	Sequence 50858, A
24	524	24.7	421	14	US-10-067-443-4	Sequence 4, Appli
25	524	24.7	421	14	US-10-067-443-28	Sequence 28, Appli
26	524	24.7	421	15	US-10-649-273-4	Sequence 4, Appli
27	524	24.7	421	15	US-10-649-273-28	Sequence 28, Appli
28	524	24.7	421	15	US-10-651-722-4	Sequence 4, Appli
29	524	24.7	421	15	US-10-651-722-28	Sequence 28, Appli
30	511.5	24.1	347	14	US-10-012-140-24	Sequence 24, Appli
31	502	23.6	348	15	US-10-282-122A-63156	Sequence 63156, A
32	494.5	23.3	343	15	US-10-282-122A-67227	Sequence 67227, A
33	492.5	23.2	341	9	US-09-815-242-11798	Sequence 11798, A
34	492.5	23.2	341	15	US-10-282-122A-66200	Sequence 66200, A
35	492.5	23.2	341	18	US-10-958-216-50	Sequence 50, Appli
36	490.5	23.1	341	18	US-10-958-216-52	Sequence 52, Appli
37	489.5	23.0	335	15	US-10-282-122A-55404	Sequence 55404, A
38	488.5	23.0	361	16	US-10-425-115-352095	Sequence 352095, A
39	484.5	22.8	337	9	US-09-815-242-10304	Sequence 10304, A
40	484.5	22.8	337	15	US-10-282-122A-56695	Sequence 56695, A
41	484.5	22.8	337	15	US-10-282-122A-75485	Sequence 75485, A
42	483.5	22.8	337	15	US-10-282-122A-78161	Sequence 78161, A
43	482.5	22.7	340	15	US-10-282-122A-68438	Sequence 68438, A
44	481.5	22.7	337	9	US-09-815-242-113780	Sequence 113780, A
45	475.5	22.4	342	9	US-09-815-242-11043	Sequence 11043, A

ALIGNMENTS

RESULT 1  
US-10-067-443-2  
; Sequence 2, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 414  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-067-443-2

Query Match 100.0%; Score 2125; DB 14; Length 414;  
Best Local Similarity 100.0%; Pred. No. 1.7e-199;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MLITKTAGVFPKSKRKRYEFLRSPNFRGTLFLKIVYLGITSCDDTPAAVDETGNV 60  
DB 1 MLITKTAGVFPKSKRKRYEFLRSPNFRGTLFLKIVYLGITSCDDTPAAVDETGNV 60  
QY 61 LGEAHSOTEVTLKNGVIFPPAQQLRHRIQRIYQVSAAGVSPSDSAITATIKPL 120  
DB 61 LGEAHSOTEVTLKNGVIFPPAQQLRHRIQRIYQVSAAGVSPSDSAITATIKPL 120

QY 121 ALSIGVLSFSLQVGLKPKPFIPIHMEAHALITRLTNKVEFPFVLLISGHCILALV 180  
DB 121 ALSIGVLSFSLQVGLKPKPFIPIHMEAHALITRLTNKVEFPFVLLISGHCILALV 180  
QY 181 QGVSDPFLILGKSLDIPAGDMLDKVARRLSLIKHPECSTMSGKAI EHLAKQGNRHFHDIK 240  
DB 181 QGVSDPFLILGKSLDIPAGDMLDKVARRLSLIKHPECSTMSGKAI EHLAKQGNRHFHDIK 240  
QY 241 PPLHAKKCDPSFTGLQHVTDKTIIMKKEKEBEGIEKGQILSSAADIAATVOHTMACHLVXR 300  
DB 241 PPLHAKKCDPSFTGLQHVTDKTIIMKKEKEBEGIEKGQILSSAADIAATVOHTMACHLVXR 300  
QY 301 THRAILFCQKRDLLPQNNNAVLVASGVSANFYIRALBELITNATQCTLLCPPRLCTDNG 360  
DB 301 THRAILFCQKRDLLPQNNNAVLVASGVSANFYIRALBELITNATQCTLLCPPRLCTDNG 360  
QY 361 IMIANGIERLRAGLGIHIDIEGIRYBPCKPLGVDISKVGEASIKVPOLKMEI 414  
DB 361 IMIANGIERLRAGLGIHIDIEGIRYBPCKPLGVDISKVGEASIKVPOLKMEI 414

RESULT 2  
US-10-649-273-2  
; Sequence 2, Application US/10649273  
; Publication No. US20040043407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649,273  
; PRIOR FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 2  
; LENGTH: 414  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-649-273-2

Query Match 100.0%; Score 2125; DB 15; Length 414;  
Best Local Similarity 100.0%; Pred. No. 1.7e-199;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MLITTTAGVFFPKSKRYEFLRSFNPHRGTLFLHKIVLGITETSCDDTAAAVDETGNV 60  
DB 1 MLITTTAGVFFPKSKRYEFLRSFNPHRGTLFLHKIVLGITETSCDDTAAAVDETGNV 60  
QY 61 LGEAHSOTEVHLKKTGGIVPPAAQOLHRENIQRIVOEALSASGVSPSDLSAATITIRGL 120  
DB 61 LGEAHSOTEVHLKKTGGIVPPAAQOLHRENIQRIVOEALSASGVSPSDLSAATITIRGL 120  
QY 121 ALSIGVLSFSLQVGLKPKPFIPIHMEAHALITRLTNKVEFPFVLLISGHCILALV 180  
DB 121 ALSIGVLSFSLQVGLKPKPFIPIHMEAHALITRLTNKVEFPFVLLISGHCILALV 180  
QY 181 QGVSDPFLILGKSLDIPAGDMLDKVARRLSLIKHPECSTMSGKAI EHLAKQGNRHFHDIK 240  
DB 181 QGVSDPFLILGKSLDIPAGDMLDKVARRLSLIKHPECSTMSGKAI EHLAKQGNRHFHDIK 240  
QY 241 PPLHAKKCDPSFTGLQHVTDKTIIMKKEKEBEGIEKGQILSSAADIAATVOHTMACHLVXR 300  
DB 241 PPLHAKKCDPSFTGLQHVTDKTIIMKKEKEBEGIEKGQILSSAADIAATVOHTMACHLVXR 300  
QY 301 THRAILFCQKRDLLPQNNNAVLVASGVSANFYIRALBELITNATQCTLLCPPRLCTDNG 360  
DB 301 THRAILFCQKRDLLPQNNNAVLVASGVSANFYIRALBELITNATQCTLLCPPRLCTDNG 360

QY 361 IMIANGIERLRAGLGIHIDIEGIRYBPCKPLGVDISKVGEASIKVPOLKMEI 414  
DB 361 IMIANGIERLRAGLGIHIDIEGIRYBPCKPLGVDISKVGEASIKVPOLKMEI 414

RESULT 3  
US-10-651-722-2  
; Sequence 2, Application US/10651722  
; Publication No. US20040048302A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 DIV  
; CURRENT APPLICATION NUMBER: US/10/651,722  
; PRIOR FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 2  
; LENGTH: 414  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-651-722-2

Query Match 100.0%; Score 2125; DB 15; Length 414;  
Best Local Similarity 100.0%; Pred. No. 1.7e-199;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITTTAGVFFPKSKRYEFLRSFNPHRGTLFLHKIVLGITETSCDDTAAAVDETGNV 60  
DB 1 MLITTTAGVFFPKSKRYEFLRSFNPHRGTLFLHKIVLGITETSCDDTAAAVDETGNV 60  
QY 61 LGEAHSOTEVHLKKTGGIVPPAAQOLHRENIQRIVOEALSASGVSPSDLSAATITIRGL 120  
DB 61 LGEAHSOTEVHLKKTGGIVPPAAQOLHRENIQRIVOEALSASGVSPSDLSAATITIRGL 120  
QY 121 ALSIGVLSFSLQVGLKPKPFIPIHMEAHALITRLTNKVEFPFVLLISGHCILALV 180  
DB 121 ALSIGVLSFSLQVGLKPKPFIPIHMEAHALITRLTNKVEFPFVLLISGHCILALV 180  
QY 181 QGVSDPFLILGKSLDIPAGDMLDKVARRLSLIKHPECSTMSGKAI EHLAKQGNRHFHDIK 240  
DB 181 QGVSDPFLILGKSLDIPAGDMLDKVARRLSLIKHPECSTMSGKAI EHLAKQGNRHFHDIK 240  
QY 241 PPLHAKKCDPSFTGLQHVTDKTIIMKKEKEBEGIEKGQILSSAADIAATVOHTMACHLVXR 300  
DB 241 PPLHAKKCDPSFTGLQHVTDKTIIMKKEKEBEGIEKGQILSSAADIAATVOHTMACHLVXR 300  
QY 301 THRAILFCQKRDLLPQNNNAVLVASGVSANFYIRALBELITNATQCTLLCPPRLCTDNG 360  
DB 301 THRAILFCQKRDLLPQNNNAVLVASGVSANFYIRALBELITNATQCTLLCPPRLCTDNG 360  
QY 361 IMIANGIERLRAGLGIHIDIEGIRYBPCKPLGVDISKVGEASIKVPOLKMEI 414  
DB 361 IMIANGIERLRAGLGIHIDIEGIRYBPCKPLGVDISKVGEASIKVPOLKMEI 414

RESULT 4  
US-10-480-988-8  
; Sequence 8, Application US/10480988  
; Publication No. US20050069877A1  
; GENERAL INFORMATION:  
; APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;  
; APPLICANT: SWARNAKAR, Anita; HAFALITA, April J.A.;  
; APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;  
; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;  
; APPLICANT: HONCHELL, Cynthia D.; NGUYEN, Damien B.;

```

; APPLICANT: LU, Duang Aina M.; LEE Ernestine A.;
; APPLICANT: YUE, Henry; FORSYTHE, Ian J.;
; APPLICANT: BARROSO, Ines; RAMKUMAR, Jayalaxmi;
; APPLICANT: GRIFFIN, Jennifer A.; LI, Joana X.;
; APPLICANT: YANG, Junning; THANGAVELU, Kavitha;
; APPLICANT: GIETZEN, Kimberley J.; DING, Li;
; APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.;
; APPLICANT: YAO, Monique G.; CHMELA, Narinder K.;
; APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
; APPLICANT: LEE, Sally; BECHA, Shanya D.;
; APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;
; APPLICANT: ELLIOTT, Vicki S.; LUO, Wen;
; APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;
; APPLICANT: LU, Yan; ZEBARUDIAN, Yeganeh
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PF-1040 USN
; CURRENT APPLICATION NUMBER: US/10/480, 988
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: PCT/US02/19360
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/300, 508
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/303, 445
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/305, 405
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/311, 442
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 60/314, 821
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/315, 992
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 60/378, 205
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ. ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7632424CD1
; US-10-480-988-8

```

Query Match	100.0%	Score 2125	DB 17	length 414
Best Local Similarity	100.0%	Pred. No. 1.7e-199		
Matches 414	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MLILTKTAGVEFKKSKRKVYEFLLSFNPHPTGLFLHKILVIGIETSCDDTPAAAVDEETGV	60	
Db	1	MLILTKTAGVEFKSKRKVYEFLLSFNPHPTGLFLHKILVIGIETSCDDTPAAAVDEETGV	60	
QY	61	LGSAIHSOTBEVHLKTGGIVPPAAQQLHRENTORIYQEAISASGVSPSDLSAIAITIKQGL	120	
Db	61	LGSAIHSOTBEVHLKTGGIVPPAAQQLHRENTORIYQEAISASGVSPSDLSAIAITIKQGL	120	
QY	121	ALSLGVGSESLQIVGQIKKPPILPHHBAHALTRLTNKPVEPPLVLLISGHCILALY	180	
Db	121	ALSLGVGSESLQIVGQIKKPPILPHHBAHALTRLTNKPVEPPLVLLISGHCILALY	180	
QY	181	QGVSDPFLILGSLDIAPGMDLKVARRSLILGHPECSIMSGGKALEHLAKQGNRHPDIK	240	
Db	181	QGVSDPFLILGSLDIAPGMDLKVARRSLILGHPECSIMSGGKALEHLAKQGNRHPDIK	240	
QY	241	PPLHAAKNCDFSTGLQHVTDKILMKKEKEBEGIKQQLSSAADIAATVQHTMACHLYKR	300	
Db	241	PPLHAAKNCDFSTGLQHVTDKILMKKEKEBEGIKQQLSSAADIAATVQHTMACHLYKR	300	
QY	301	THRAILTECKQDILLPQNNATVYAGGASNSPTIRALALSLTAAITQCTLLCAPPKALYUNG	360	
Db	301	THRAILTECKQDILLPQNNATVYAGGASNSPTIRALALSLTAAITQCTLLCAPPKALYUNG	360	

```

Oy      361  |||||MM|MMNGIERLRAGAGLGHDIHDEIGIRYEPKCPGLGYDISEKVEBASIKVPOLKMEI 414
Db      361  |||||IM|MMNGIERLRAGAGLGHDIHDEIGIRYEPKCPGLGYDISEKVEBASIKVPOLKMEI 414

RESULT 5
US-10-067-443-19
; Sequence 19, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
; TITLE OF INVENTION: SPINAL CORD, NP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRt
; ORGANISM: homo sapiens
; US-10-067-443-19

```

```

Query Match 98.4%; Score 2090.5; DB 14; Length 439;
Best Local Similarity 99.8%; Pred. No. 4,66-196;
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY      1 MLIIITKAGVFPKRSKRKYVEFLRSFNFHPTLFLAKIVLGIETSCDDTAAYVDEGTNV 60
DB      1 MLIIITKAGVFPKRSKRKYVEFLRSFNFHPTLFLAKIVLGIETSCDDTAAYVDEGTNV 60
QY      61 LGEAHSOTVEHLKTGGIVPPAAQOLHRENTQRIVOEALSSAGVSPSDLSAIAATTIKPGL 120
DB      61 LGEAHSOTVEHLKTGGIVPPAAQOLHRENTQRIVOEALSSAGVSPSDLSAIAATTIKPGL 120
QY      121 ALSIVGSLSPSLQIVLQVGLKKPFIPIHMEAHALTRLTNNKVEPFPVLIIISGCHLALV 180
DB      121 ALSIVGSLSPSLQIVLQVGLKKPFIPIHMEAHALTRLTNNKVEPFPVLIIISGCHLALV 180
QY      181 QGVSDPFLILKSLDIAAGPMLDKVARSLSLIKHEPCSTMSGKAI EHLAKGNRPHPIK 240
DB      181 QGVSDPFLILKSLDIAAGPMLDKVARSLSLIKHEPCSTMSGKAI EHLAKGNRPHPIK 240
QY      241 PPLHAKNCPFSFTGLQHTVTDKIIMKKEKEGCI-----EK 275
DB      241 PPLHAKNCPFSFTGLQHTVTDKIIMKKEKEGCI-----EK 275
QY      276 GQIISSAADIATVQHTMACHLVKTRTRAILFCRQDILLPONNAVVASGVASNFYIRR 335
DB      276 GQIISSAADIATVQHTMACHLVKTRTRAILFCRQDILLPONNAVVASGVASNFYIRR 335
QY      301 GQIISSAADIATVQHTMACHLVKTRTRAILFCRQDILLPONNAVVASGVASNFYIRR 360
DB      301 GQIISSAADIATVQHTMACHLVKTRTRAILFCRQDILLPONNAVVASGVASNFYIRR 360
QY      336 ALELTMAQCTLLCPPRCTDNGIMIANNGIRLAAAGGILHIDIEIRREPCPLGYD 395
DB      361 ALELTMAQCTLLCPPRCTDNGIMIANNGIRLAAAGGILHIDIEIRREPCPLGYD 420
QY      396 ISKEVGASIKVPOLKMEI 414
DB      421 ISKEVGASIKVPOLKMEI 439

RESULT 6
US-10-649-273-19
; Sequence 19, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT

```

```

; CURRENT APPLICATION NUMBER: US/10/649, 273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-649-273-19

```

```

Query Match      98.4%; Score 2090.5; DB 15; Length 439;
Best Local Similarity 93.8%; Pred. No. 4.6e-196;
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

```

```

QY 1 MLITKTAGVFFPKSKRYEFLRSFNHFGTLFLHKIVLGIEISCDPTAAAVDETVGV 60
DB 1 MLITKTAGVFFPKSKRYEFLRSFNHFGTLFLHKIVLGIEISCDPTAAAVDETVGV 60
QY 61 LGEAHSQTEVHLKTGGIVPPAQQIHRNRIQIVQEAALSASGVSPSDLSAIAITTIKPGI 120
DB 61 LGEAHSQTEVHLKTGGIVPPAQQIHRNRIQIVQEAALSASGVSPSDLSAIAITTIKPGI 120
QY 121 ALISGVLSFSLQVGLQKKPPIPIHMEAHALTRLTNKKVEPPLVLLISGHCILALV 180
DB 121 ALISGVLSFSLQVGLQKKPPIPIHMEAHALTRLTNKKVEPPLVLLISGHCILALV 180
QY 181 QGVDFLLGKSLDIAFGMDLVKARRSLIKHPECSTMSGKAIHLAKQNRFFFDIK 240
DB 181 QGVDFLLGKSLDIAFGMDLVKARRSLIKHPECSTMSGKAIHLAKQNRFFFDIK 240
QY 241 PPLHAKNCDPFTGLQHTVTDKIIMKKEKEGFI-----EK 275
DB 241 PPLHAKNCDPFTGLQHTVTDKIIMKKEKEGFI-----EK 275
QY 276 GQILSSADIAATVQHTMACHLVKTRHAILFCRKORDLLPQNNAVLVASGVASNFYIR 335
DB 301 GQILSSADIAATVQHTMACHLVKTRHAILFCRKORDLLPQNNAVLVASGVASNFYIR 360
QY 336 ALBITLNAQTCTLCPPLCTDNGIMIAMNGIERLRAGLIGLHIDIEGIRYEPKCPGLVD 395
DB 361 ALBITLNAQTCTLCPPLCTDNGIMIAMNGIERLRAGLIGLHIDIEGIRYEPKCPGLVD 420
QY 396 ISKEVGEASIKVPOLKMEI 414
DB 421 ISKEVGEASIKVPOLKMEI 439

```

```

RESULT 7
US-10-651-722-19
; Sequence 19, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 19
; LENGTH: 439

```

```

; TYPE: PRT
; ORGANISM: homo sapiens
US-10-651-722-19

```

```

Query Match      98.4%; Score 2090.5; DB 15; Length 439;
Best Local Similarity 93.8%; Pred. No. 4.6e-196;
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

```

```

QY 1 MLITKTAGVFFPKSKRYEFLRSFNHFGTLFLHKIVLGIEISCDPTAAAVDETVGV 60
DB 1 MLITKTAGVFFPKSKRYEFLRSFNHFGTLFLHKIVLGIEISCDPTAAAVDETVGV 60
QY 61 LGEAHSQTEVHLKTGGIVPPAQQIHRNRIQIVQEAALSASGVSPSDLSAIAITTIKPGI 120
DB 61 LGEAHSQTEVHLKTGGIVPPAQQIHRNRIQIVQEAALSASGVSPSDLSAIAITTIKPGI 120
QY 121 ALISGVLSFSLQVGLQKKPPIPIHMEAHALTRLTNKKVEPPLVLLISGHCILALV 180
DB 121 ALISGVLSFSLQVGLQKKPPIPIHMEAHALTRLTNKKVEPPLVLLISGHCILALV 180
QY 181 QGVDFLLGKSLDIAFGMDLVKARRSLIKHPECSTMSGKAIHLAKQNRFFFDIK 240
DB 181 QGVDFLLGKSLDIAFGMDLVKARRSLIKHPECSTMSGKAIHLAKQNRFFFDIK 240
QY 241 PPLHAKNCDPFTGLQHTVTDKIIMKKEKEGFI-----EK 275
DB 241 PPLHAKNCDPFTGLQHTVTDKIIMKKEKEGFI-----EK 275
QY 276 GQILSSADIAATVQHTMACHLVKTRHAILFCRKORDLLPQNNAVLVASGVASNFYIR 335
DB 301 GQILSSADIAATVQHTMACHLVKTRHAILFCRKORDLLPQNNAVLVASGVASNFYIR 360
QY 336 ALBITLNAQTCTLCPPLCTDNGIMIAMNGIERLRAGLIGLHIDIEGIRYEPKCPGLVD 395
DB 361 ALBITLNAQTCTLCPPLCTDNGIMIAMNGIERLRAGLIGLHIDIEGIRYEPKCPGLVD 420
QY 396 ISKEVGEASIKVPOLKMEI 414
DB 421 ISKEVGEASIKVPOLKMEI 439

```

```

RESULT 8
US-10-012-140-5
; Sequence 5, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kappelner-Libermann, Rosana
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; FILE REFERENCE: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; TITLE OF INVENTION: THERBOF
; CURRENT APPLICATION NUMBER: US/10/012,140
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772-yes
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-140-5

```

```

Query Match      98.3%; Score 2088; DB 14; Length 414;
Best Local Similarity 98.3%; Pred. No. 7.5e-196;
Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 MLITKTAGVFPKSKRVYEFIRSFNHPGTLFLHKTIVLGIETSCDDTAAAVDDEGNV 60  
DB 1 MLITKTAGVFPKSKRVYEFIRSFNHPGTLFLHKTIVLGIETSCDDTAAAVDDEGNV 60  
QY 61 LGEAHSQTEVHLKTGCIYPPAAQOLHRENIQRIYQVQALASASVSPSDLSAATTIKPGL 120  
DB 61 LGEAHSQTEVHLKTGCIYPPAAQOLHRENIQRIYQVQALASASVSPSDLSAATTIKPGL 120  
QY 121 ALSIGVGSFSLQVLGQKKPPIPIHMEAHALTRLTNKNVFPFVLLISGCHLLALV 180  
DB 121 ALSIGVGSFSLQVLGQKKPPIPIHMEAHALTRLTNKNVFPFVLLISGCHLLALV 180  
QY 181 QGVSDPFLGKSLDIAPGMDLVKVARRLSLIKHPECSITSGGKAIEHLAKQGNRPFHFDIK 240  
DB 181 QGVSDPFLGKSLDIAPGMDLVKVARRLSLIKHPECSITSGGKAIEHLAKQGNRPFHFDIK 240  
QY 241 PPLHAKNCDPFTGLQHTVTDKIIMKKEBEGIEKQIILSSADIAATVQHTMACHLVKR 300  
DB 241 PPLHAKNCDPFTGLQHTVTDKIIMKKEBEGIEKQIILSSADIAATVQHTMACHLVKR 300  
QY 301 THRALLFCQRDLPPONNAVLVASGVASNFYIRRALLETITNATQCTLLCPPRLCTDNG 360  
DB 301 THRALLFCQRDLPPONNAVLVASGVASNFYIRRALLETITNATQCTLLCPPRLCTDNG 360  
QY 361 IMIANGIERLRAGIGIHDIEGIRYEPKCPGVDISKEVGEASIKVPOLKMEI 414  
DB 361 IMIANGIERLRAGIGIHDIEGIRYEPKCPGVDISKEVGEASIKVPOLKMEI 414

RESULT 9  
US-10-094-749-2039  
; Sequence 2039, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOKYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; PRIOR FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2039  
; LENGTH: 364  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-094-749-2039

Query Match 86.8%; Score 1845; DB 15; Length 364;  
Best Local Similarity 93.2%; Pred. No. 4, 5e-172;  
Matches 361; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MLITKTAGVFPKSKRVYEFIRSFNHPGTLFLHKTIVLGIETSCDDTAAAVDDEGNV 60

DB 1 MLITKTAGVFPKSKRVYEFIRSFNHPGTLFLHKTIVLGIETSCDDTAAAVDDEGNV 60  
QY 61 LGEAHSQTEVHLKTGCIYPPAAQOLHRENIQRIYQVQALASASVSPSDLSAATTIKPGL 120  
DB 61 LGEAHSQTEVHLKTGCIYPPAAQOLHRENIQRIYQVQALASASVSPSDLSAATTIKPGL 120  
QY 121 ALSIGVGSFSLQVLGQKKPPIPIHMEAHALTRLTNKNVFPFVLLISGCHLLALV 180  
DB 121 ALSIGVGSFSLQVLGQKKPPIPIHMEAHALTRLTNKNVFPFVLLISGCHLLALV 180  
QY 181 QGVSDPFLGKSLDIAPGMDLVKVARRLSLIKHPECSITSGGKAIEHLAKQGNRPFHFDIK 240  
DB 181 QGVSDPFLGKSLDIAPGMDLVKVARRLSLIKHPECSITSGGKAIEHLAKQGNRPFHFDIK 240  
QY 241 PPLHAKNCDPFTGLQHTVTDKIIMKKEBEGIEKQIILSSADIAATVQHTMACHLVKR 300  
DB 241 PPLHAKNCDPFTGLQHTVTDKIIMKKEBEGIEKQIILSSADIAATVQHTMACHLVKR 300  
QY 301 THRALLFCQRDLPPONNAVLVASGVASNFYIRRALLETITNATQCTLLCPPRLCTDNG 360  
DB 301 THRALLFCQRDLPPONNAVLVASGVASNFYIRRALLETITNATQCTLLCPPRLCTDNG 360  
QY 361 IMIA 364  
DB 361 IMIA 364

RESULT 10  
US-10-067-443-22  
; Sequence 22, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 22  
; LENGTH: 267  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-067-443-22

Query Match 65.2%; Score 1385; DB 14; Length 267;  
Best Local Similarity 100.0%; Pred. No. 4, 2e-127;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 148 MEAHALTRLTNKNVFPFVLLISGCHLLAVQGVSDPFLGKSLDIAPGMDLVKVAR 207  
DB 1 MEAHALTRLTNKNVFPFVLLISGCHLLAVQGVSDPFLGKSLDIAPGMDLVKVAR 60  
QY 208 LSLIKHPECSITSGGKAIEHLAKQGNRPFHFDIKPPLHAKNCDPFTGLQHTVTDKIIMK 267  
DB 208 LSLIKHPECSITSGGKAIEHLAKQGNRPFHFDIKPPLHAKNCDPFTGLQHTVTDKIIMK 120  
QY 268 EKEBEGIEKQIILSSADIAATVQHTMACHLVKRTTHRALLFCQRDLPPONNAVLVASGV 327  
DB 268 EKEBEGIEKQIILSSADIAATVQHTMACHLVKRTTHRALLFCQRDLPPONNAVLVASGV 180  
QY 121 EKEBEGIEKQIILSSADIAATVQHTMACHLVKRTTHRALLFCQRDLPPONNAVLVASGV 180  
DB 121 EKEBEGIEKQIILSSADIAATVQHTMACHLVKRTTHRALLFCQRDLPPONNAVLVASGV 180  
QY 328 ASNFYIRRALLETITNATQCTLLCPPRLCTDNGIMIANNGIERLRAGIGIHDIEGIRY 387  
DB 328 ASNFYIRRALLETITNATQCTLLCPPRLCTDNGIMIANNGIERLRAGIGIHDIEGIRY 240  
QY 368 PKCPLGVDSKEVGEASIKVPOLKMEI 414  
DB 368 PKCPLGVDSKEVGEASIKVPOLKMEI 414

Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

## RESULT 11

US-10-649-273-22  
; Sequence 22, Application US/10649273  
; Publication No. US20040043407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649,273  
; PRIOR FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 22  
; LENGTH: 267  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-649-273-22

Query Match 65.2%; Score 1385; DB 15; Length 267;  
Best Local Similarity 100.0%; Pred. No. 4.2e-127;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 MEAHALITRLTNKVEFPFLVLLISGGHCLALVGVSDFLLGKSLDIAPGMDLKVARR 207  
Db 1 MEAHALITRLTNKVEFPFLVLLISGGHCLALVGVSDFLLGKSLDIAPGMDLKVARR 60  
Qy 208 LSLIKHPECSTMSGGKAIEHLAKOGRFFFDIKPPLHAKNCDPFSFTGQHTVDKIMKK 267  
Db 61 LSLIKHPECSTMSGGKAIEHLAKOGRFFFDIKPPLHAKNCDPFSFTGQHTVDKIMKK 120  
Qy 268 EKEBGEIKGQILSSADIAATVQHTMACHLVKTRTRAILFCQKRDLLPQNNAVLVASGV 327  
Db 121 EKEBGEIKGQILSSADIAATVQHTMACHLVKTRTRAILFCQKRDLLPQNNAVLVASGV 180  
Qy 328 ASNFYIRRALEILTNAOTCTLLCPPRLCTDNGIMIANNGIERLRAGLGIHDIISIRYE 387  
Db 181 ASNFYIRRALEILTNAOTCTLLCPPRLCTDNGIMIANNGIERLRAGLGIHDIISIRYE 240  
Qy 388 PKCPLGVDISKVEGASIKVPOLKMEI 414  
Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

## RESULT 12

US-10-651-722-22  
; Sequence 22, Application US/10651722  
; Publication No. US20040048302A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1  
; FILE REFERENCE: D0073 DIV  
; CURRENT APPLICATION NUMBER: US/10/651,722  
; CURRENT FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 22  
; LENGTH: 267  
; TYPE: PRT

; ORGANISM: homo sapiens  
US-10-651-722-22

Query Match 65.2%; Score 1385; DB 15; Length 267;  
Best Local Similarity 100.0%; Pred. No. 4.2e-127;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 MEAHALITRLTNKVEFPFLVLLISGGHCLALVGVSDFLLGKSLDIAPGMDLKVARR 207  
Db 1 MEAHALITRLTNKVEFPFLVLLISGGHCLALVGVSDFLLGKSLDIAPGMDLKVARR 60  
Qy 208 LSLIKHPECSTMSGGKAIEHLAKOGRFFFDIKPPLHAKNCDPFSFTGQHTVDKIMKK 267  
Db 61 LSLIKHPECSTMSGGKAIEHLAKOGRFFFDIKPPLHAKNCDPFSFTGQHTVDKIMKK 120  
Qy 268 EKEBGEIKGQILSSADIAATVQHTMACHLVKTRTRAILFCQKRDLLPQNNAVLVASGV 327  
Db 121 EKEBGEIKGQILSSADIAATVQHTMACHLVKTRTRAILFCQKRDLLPQNNAVLVASGV 180  
Qy 328 ASNFYIRRALEILTNAOTCTLLCPPRLCTDNGIMIANNGIERLRAGLGIHDIISIRYE 387  
Db 181 ASNFYIRRALEILTNAOTCTLLCPPRLCTDNGIMIANNGIERLRAGLGIHDIISIRYE 240  
Qy 388 PKCPLGVDISKVEGASIKVPOLKMEI 414  
Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

## RESULT 13

US-11-097-143-34191  
; Sequence 34191, Application US/11097143  
; Publication No. US20050208558A1  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
; FILE REFERENCE: C1000728  
; CURRENT APPLICATION NUMBER: US/11/097,143  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: 60/157,832  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: 60/160,191  
; PRIOR FILING DATE: 1998-10-19  
; PRIOR APPLICATION NUMBER: 60/161,932  
; PRIOR FILING DATE: 1999-10-28  
; PRIOR APPLICATION NUMBER: 60/164,769  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/173,383  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: 60/175,693  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/184,831  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/191,637  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 43008  
; SOFTWARE: FastSeq for windows Version 4.0  
; SEQ ID NO 34191  
; LENGTH: 409  
; TYPE: PRT  
; ORGANISM: DROSOPHILA  
US-11-097-143-34191

Query Match 33.6%; Score 714.5; DB 20; Length 409;  
Best Local Similarity 42.0%; Pred. No. 5.7e-61;  
Matches 156; Conservative 60; Mismatches 140; Indels 15; Gaps 6;

Qy 39 VLGIEFTSCDDTAADVDEGVNIGALIHSGTEVHLTKGTGIVPPAQOLHRENIQRTVQEA 98  
Db 27 VLGIEFTSCDDTAADVDEGVNIGALIHSGTEVHLTKGTGIVPPAQOLHRENIQRTVQEA 86



QY 99 LSASGVSFSDLSAIAATTIKPGLAISGVLSQVLQVQQLKKRPIPIHMEHALTIRUT 158  
DB 87 MEAAQLKPDQLTAIAVTRPGLPLSLVGVRFARHARLQKPLPIVHMEHALQARRE 146  
QY 159 N--KVEPFLVLLISGGHCLLALVQGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPEC 216  
DB 147 HPEIGVFPFLCLLASGGHCLLVANGPGRLLTIGTLDDBEADDKGRRLRIHLPEY 206  
QY 217 STMSGKAIEHLAK--QGNRFHFDIKRPLHAKNCDPSTGLQHTVDKIMKEKEGIEK 275  
DB 207 RLWNGRAIEHAQAQADPLAVEPFLPLAQQRNCNCFSTAGIKNNSFRRAIARERARPTP 266  
QY 276 GQILSSADIAATVQHTMACHLVKRTTRAILFC--KQDPLPQNNAVLVASGVASNFYI 333  
DB 267 DGVISNYGDFCAGLRVSRLMHTORAEICLLPHRQLFQDTPPTLVMSGGVANNDAI 326  
QY 334 RBAEILTNATQCTLLCPPLCTDNGIMIANNGIERLRAGILHDE--GIRYEPKCL 392  
DB 327 YANIEHLAAQYGCSPFRPSKRYCSDNGVMIAMHGEVL-----LQDKEASTRYD-- 377  
QY 393 GVDISKEVGEA 403  
DB 378 SIDIQSAGFA 388  
RESULT 14  
US-10-424-599-209259  
; Sequence 209259, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 209259  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_3098C.1.pep  
US-10-424-599-209259  
Query Match 32.1%; Score 681.5; DB 15; Length 445;  
Best Local Similarity 43.1%; Pred. No. 1,1e-57;  
Matches 155; Conservative 54; Mismatches 126; Indels 25; Gaps 6;  
QY 38 IVLGIEFSCDDTAAAVDETGNVLGEALHSQTEVHLKTGGIVPPAAQQLHRENIQRIVOE 97  
DB 55 VVLGIEFSCDDTAAAVVDSDEILSQVVSQADLLAKGVAPKMAEASHKVIDQVVOE 114  
QY 98 ALSASGVSPDLSAIAATTIKPGLAISGVLSQVLQVQQLKKRPIPIHMEHALTIRL 157  
DB 115 ALDVAIVTEKDLTAIAVTRPGLPLSLVGVRFARHARLQKPLPIVHMEHALQARRE 174  
QY 158 TNK--VEPFLVLLISGGHCLLALVQGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPEC 216  
DB 175 IEKDLOPFPFLMALISGGHCLLALVQGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPEC 231  
QY 217 STMSGKAIEHLAK--QGNRFHFDIKRPLHAKNCDPSTGLQHTVDKIMKEKEGIEK 271  
DB 232 --RSGGPAIEHLAK--QGNRFHFDIKRPLHAKNCDPSTGLQHTVDKIMKEKEGIEK 289  
QY 272 GIEKQILSSADIAATVQHTMACHLVKRTTRAILFC--KQDPLPQNNAVLVASGVASNF 331  
DB 290 SASNGRDL--SPADIAASFQRIAVLHLBERCERAIQWALKMPSIRH--LVVSGGVASNQ 345  
QY 332 YIRBAEILTNATQCTLLCPPLCTDNGIMIANNGIERLRAGILHDEGIRYEPKCP 391

DB 346 YIRBAEILTNATQCTLLCPPLCTDNGIMIANNGIERLRAGILHDEGIRYEPKCP 395  
RESULT 15  
US-10-067-443-3  
; Sequence 3, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 3  
; LENGTH: 463  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-10-067-443-3  
Query Match 29.8%; Score 634; DB 14; Length 463;  
Best Local Similarity 37.9%; Pred. No. 5.6e-53;  
Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;  
QY 38 IVLGIEFSCDDTAAAVDETGNVLGEALHSQTEVHLKTGGIVPPAAQQLHRENIQRIVOE 97  
DB 85 VVLGIEFSCDDTAAAVVDSDEILSQVVSQADLLAKGVAPKMAEASHKVIDQVVOE 141  
QY 98 ALSASGVSPDLSAIAATTIKPGLAISGVLSQVLQVQQLKKRPIPIHMEHALTIRL 157  
DB 142 ALDKANLTKEDLSAIAVTRPGLPLSLVGVRFARHARLQKPLPIVHMEHALQARRE 201  
QY 158 T--NKVEPFLVLLISGGHCLLALVQGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPEC 216  
DB 202 VQELSPFPFLMALISGGHCLLALVQGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPEC 258  
QY 217 STMSGKAIEHLAK--QGNRFHFDIKRPLHAKNCDPSTGLQHTVDKIMKEKEGIEK 276  
DB 259 --RSGGPAIEHLAK--QGNRFHFDIKRPLHAKNCDPSTGLQHTVDKIMKEKEGIEK 308  
QY 277 QILSSADIAATVQHTMACHLVKRTTRAILFC--KQDPLPQNNAVLVASGVASNFYIRRA 336  
DB 309 --IRNRADIAASFQRIAVLHLBERCERAIQWALKMPSIRH--LVVSGGVASNQ 363  
QY 337 IEILTNATQCTLLCPPLCTDNGIMIANNGIERLRAGILHDEGIRYEPKCP 391  
DB 364 LNNIIVENKRLAKLVCPPLCTDNGIMIANNGIERLRAGILHDEGIRYEPKCP 413  
QY 388 -----PKCPPLGVDISKEVGEA 403  
DB 414 DYVVDLPRPWPFGEEYAKGRSEA 436

Search completed: November 10, 2005, 16:58:50  
Job time : 112.35 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:02:11 ; Search time 111.6 Seconds  
(without alignment)  
1434.756 Million cell updates/sec

Title: US-10-649-273-2  
Sequence: 1 MLITRTAGVFFKPSKRVY.....DISKEVGASIKVPLKMEI 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1980s:\*  
3: geneseqp2000s:\*  
4: geneseqp2000s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	414	5	ABG96478 Novel hum
2	2125	100.0	414	5	ABG05481 Human O-s
3	2125	100.0	414	6	ABJ26654 Human pro
4	2088	98.3	414	5	AAE29234 Human gly
5	2088	98.3	414	6	ABG71161 Novel hum
6	2088	98.3	414	6	ABU09569 Human gly
7	1845	86.8	364	6	ADA54471 Human pro
8	1385	65.2	267	5	ABG96487 Novel hum
9	714.5	33.6	409	4	ABG69133 Drosophil
10	659.5	31.0	439	3	AAI19286 Arabidops
11	659.5	31.0	444	3	AAI19285 Arabidops
12	634	29.8	463	3	AAV52216 Arabidops
13	549	25.8	179	5	ABG96489 Novel hum
14	524	24.7	382	6	ABU22934 Protein e
15	524	24.7	421	5	ABG96491 Novel hum
16	502	23.6	348	6	ABU35232 Protein e
17	501	23.6	350	8	ADL05040 M. catarr
18	499.5	23.5	401	7	ABO68626 Pseudomon
19	494.5	23.3	343	6	ABU39303 Protein e
20	492.5	23.2	341	4	AAU36205 Pseudomon
21	492.5	23.2	341	6	ABU38276 Protein e
22	490.5	23.2	341	7	ADG73342 P aerugin
23	489.5	23.0	335	6	ABU27480 Protein e
24	489.5	23.0	337	4	AAU34711 E. coli c
25	484.5	22.8	337	4	AAU34711 E. coli c

26	484.5	22.8	337	6	ABU47561 Protein e
27	484.5	22.8	337	6	ABU28771 Protein e
28	483.5	22.8	337	6	ABU50237 Protein e
29	482.5	22.7	340	6	ABU40514 Protein e
30	482.5	22.7	357	7	ADP06228 Bacteri
31	481.5	22.7	337	4	AAU38187
32	479.5	22.6	337	3	AAV52204
33	475.5	22.4	335	2	AAV52203
34	475.5	22.4	335	3	AAV52202
35	475.5	22.4	342	3	AAV52202
36	475.5	22.4	342	4	AAU35450
37	475.5	22.4	342	4	AAU30280
38	474.5	22.3	343	7	ABO62704
39	472.5	22.2	338	6	ABM67812
40	472.5	22.2	363	7	ADC69682
41	469	22.1	340	6	ABU24130
42	468	22.0	338	6	ABU29893
43	465.5	21.9	346	3	AAV52208
44	465.5	21.9	354	6	ABU37844
45	465	21.9	341	6	ABU40069

## ALIGNMENTS

RESULT 1  
ABG96478  
ID ABG96478 standard; protein; 414 AA.  
XX  
AC ABG96478;  
XX  
DT 11-DEC-2002 (first entry)  
XX  
DE Novel human metalloprotease MPL.  
XX  
KW Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
KW neurological disorder.  
XX  
OS Homo sapiens.  
XX  
XX W020027251-A2.  
XX  
PD 19-SEP-2002.  
XX  
PF 05-FEB-2002; 2002WO-US003353.  
XX  
PR 05-FEB-2001; 2001US-0266518P.  
XX  
PR 10-APR-2001; 2001US-0282814P.  
XX  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S;  
XX  
XX MPI: 2002-723329/78.  
XX  
XX N-PSDB; ABS76635.  
XX  
XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
XX treating, or ameliorating diseases associated with aberrant  
XX metalloprotease activity, e.g. immune, metabolic, inflammatory and  
XX neurological disorders.  
XX  
XX Claim 5; Fig 1A-C; 473pp; English.  
XX  
XX The invention describes an isolated nucleic acid molecule (I) encoding a  
XX metalloprotease (MP-1). (I) is useful for preventing, treating, or

CC ameliorating a medical condition, particularly an immune disorder, an  
CC aberrant glutamate transport or motor neuron disorder, such as  
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
CC condition. The compositions and methods are also useful for diagnosing,  
CC prognosticating, treating, ameliorating and/or treating disorders  
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
CC sequence of a metalloprotease Mpl protein  
XX

SQ Sequence 414 AA;

Query Match 100.0%; Score 2125; DB 5; Length 414;  
Best Local Similarity 100.0%; Pred. No. 2.2e-215;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITKTAGVFPKSKRYEFLRSFNFPGTLFLHKIVLGITSCDDTAAAVDETVGNV 60  
DB 1 MLITKTAGVFPKSKRYEFLRSFNFPGTLFLHKIVLGITSCDDTAAAVDETVGNV 60  
QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIQVEALASASVSPSDLSAIAITTKPGL 120  
DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIQVEALASASVSPSDLSAIAITTKPGL 120  
QY 121 ALSLGVGSFSLQVLQGLKKPFIPIHMEAHALTITLTKVPEFPFLVLLISGGHCLLAV 180  
DB 121 ALSLGVGSFSLQVLQGLKKPFIPIHMEAHALTITLTKVPEFPFLVLLISGGHCLLAV 180  
QY 181 QGVSDFLLGSLDIAFGMDLVKARSLSIKPECSYWSGGKAIIEHLAKQGRFHDIK 240  
DB 181 QGVSDFLLGSLDIAFGMDLVKARSLSIKPECSYWSGGKAIIEHLAKQGRFHDIK 240  
QY 241 PRLHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYKR 300  
DB 241 PRLHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYKR 300  
QY 301 THRALLFCQKRDLLPQNNNAVLVAGGVASNFYIRRALIELITNAOCTLLCPPRLCTDNG 360  
DB 301 THRALLFCQKRDLLPQNNNAVLVAGGVASNFYIRRALIELITNAOCTLLCPPRLCTDNG 360  
QY 361 IMIANGIERLRAGLGIHLDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414  
DB 361 IMIANGIERLRAGLGIHLDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414

## RESULT 2

AB05481  
ID AB05481 standard; protein; 414 AA.

AC ABB05481;

DT 19-APR-2002 (first entry)

XX Human O-sialoglycoproteinase-like protein SEQ ID NO:2.

XX Human; O-sialoglycoproteinase-like protein; OSGPLP; enzyme.

XX Homo sapiens.

XX CN1318550-A.

XX 24-OCT-2001.

XX 19-APR-2000; 2000CN-00106834.

XX 19-APR-2000; 2000CN-00106834.

XX

PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.

XX Mao Y, Xie Y;

XX WPI; 2002-115090/16.

XX N-PSDB; ABA93268.

PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful  
for diagnosing, preventing and treating related diseases.

PS Claim 1; Page 30-31 (Disclosure); 38pp; Chinese.

CC The present sequence represents human O-sialoglycoproteinase-like protein  
CC (OSGPlp). The present invention also describes: (1) the preparation of  
CC the OSGPlp protein; (2) applying the OSGPlp protein in diagnosis; (3) the  
CC prevention and/or treatment of related diseases; (4) utilizing the OSGPlp  
CC protein in screening its agonist, excitomotor and inhibitor and preparing  
CC an antibody against the OSGPlp protein; and (5) the use of the OSGPlp  
CC polynucleotide sequences, proteins, agonists, excitomotors, inhibitors  
CC and antibodies in treating diseases related to the abnormal OSGPlp gene  
CC and in preparing the medicine composite for the treatment

SQ Sequence 414 AA;

Query Match 100.0%; Score 2125; DB 5; Length 414;  
Best Local Similarity 100.0%; Pred. No. 2.2e-215;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITKTAGVFPKSKRYEFLRSFNFPGTLFLHKIVLGITSCDDTAAAVDETVGNV 60  
DB 1 MLITKTAGVFPKSKRYEFLRSFNFPGTLFLHKIVLGITSCDDTAAAVDETVGNV 60  
QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIQVEALASASVSPSDLSAIAITTKPGL 120  
DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIQVEALASASVSPSDLSAIAITTKPGL 120  
QY 121 ALSLGVGSFSLQVLQGLKKPFIPIHMEAHALTITLTKVPEFPFLVLLISGGHCLLAV 180  
DB 121 ALSLGVGSFSLQVLQGLKKPFIPIHMEAHALTITLTKVPEFPFLVLLISGGHCLLAV 180  
QY 181 QGVSDFLLGSLDIAFGMDLVKARSLSIKPECSYWSGGKAIIEHLAKQGRFHDIK 240  
DB 181 QGVSDFLLGSLDIAFGMDLVKARSLSIKPECSYWSGGKAIIEHLAKQGRFHDIK 240  
QY 241 PRLHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYKR 300  
DB 241 PRLHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYKR 300  
QY 301 THRALLFCQKRDLLPQNNNAVLVAGGVASNFYIRRALIELITNAOCTLLCPPRLCTDNG 360  
DB 301 THRALLFCQKRDLLPQNNNAVLVAGGVASNFYIRRALIELITNAOCTLLCPPRLCTDNG 360  
QY 361 IMIANGIERLRAGLGIHLDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414  
DB 361 IMIANGIERLRAGLGIHLDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414

## RESULT 3

ABJ26654  
ID ABJ26654 standard; protein; 414 AA.

AC ABJ26654;

DT 01-MAY-2003 (first entry)

XX Human protein modification + maintenance molecule protein SEQ ID NO:8.

XX Cystostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;  
XX cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;  
XX anticancer; hepatocytotoxic; gynaecological; antibacterial; virucide;  
XX protozoacide; antiparasitic; cell proliferative disease; PMOD;  
XX protein modification and maintenance molecule; immunogenic fragment;  
XX cancer; autoimmune; inflammatory disease; neurological disorder;



PN WO200274960-A2.  
 XX 26-SEP-2002.  
 XX 08-NOV-2001; 2001WO-US051427.  
 PF 08-NOV-2000; 2000US-0246768P.  
 PR 08-NOV-2000; 2000US-0246772P-yes  
 PR 15-NOV-2000; 2000US-0249185P.  
 XX (MILL-) MILLENNIUM PHARM INC.  
 PA  
 PI Leyby KR, Kapeller-Libermann R, Glucksmann M;  
 XX WPI: 2002-759898/82.  
 DR N-PSDB; AAD46856.  
 XX  
 PT New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,  
 PT useful for diagnosing and treating cancer, immune, cardiovascular,  
 PT hemopoietic, brain, pain, metabolic, liver or platelet disorders, and  
 PT in pharmacogenomics.  
 XX  
 PS Claim 1; Fig 8; 178pp; English.  
 XX  
 CC The present invention relates to novel 38650, 28472, 5495, 65507, 81588  
 CC or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-  
 CC protease or seven transmembrane domain (7TM) receptor family members.  
 CC Sequences of seven transmembrane domain (7TM) receptor family members.  
 CC or aberrant cellular proliferation and/or differentiation (e.g. colon or  
 CC lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid  
 CC arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,  
 CC hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,  
 CC myocardial infarction, thrombus) including endothelial cell disorders  
 CC (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain  
 CC disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),  
 CC pain and metabolic disorders (e.g. obesity), liver disorders or platelet  
 CC disorders. They are also useful in screening assays, predictive medicine  
 CC (e.g. diagnostic assays, prognostic assays, monitoring clinical trials  
 CC and pharmacogenetics) and prophylactic and therapeutic methods. The  
 CC nucleic acids may also be used in chromosome mapping, tissue typing and  
 CC forensic biology and as surrogate markers. Sequences of the invention are  
 CC also used in gene therapy. The present sequence is human glycoprotease  
 CC 28472 protein  
 XX  
 SQ Sequence 414 AA;  
 Query Match 98.3%; Score 2088; DB 5; Length 414;  
 Best Local Similarity 98.3%; Pred. No. 1.9e-211;  
 Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 361 IMIAMNGIERLRAGLGLHDIEGIRYEPKCPGLVDISKVEGBASIKVPOLMEI 414  
 DB 361 IMIAMNGIERLRAGLGLHDIEGIRYEPKCPGLVDISKVEGBASIKVPOLMEI 414  
 RESULT 5  
 ABG71161  
 ID ABG71161 standard; protein; 414 AA.  
 XX  
 AC ABG71161;  
 XX  
 DT 30-JAN-2003 (first entry)  
 DE Novel human glycoprotease 28472.  
 XX  
 KW Cancer; aberrant cell proliferation; aberrant cell differentiation;  
 KW breast cancer; ovarian cancer; prostate cancer; colon cancer;  
 KW lung cancer; immune disorder; heart disorder; cardiovascular disorder;  
 KW endothelial disorder; hemopoietic disorder; blood vessel disorder;  
 KW brain disorder; pain; metabolic disorder; liver disorder; diabetes;  
 KW platelet disorder; carcinoma; sarcoma; leukaemia; Hodgkin's disease;  
 KW autoimmune disorder; hypertension; atherosclerosis; heart failure;  
 KW myocardial infarction; ischaemic heart disease; Crohn's disease;  
 KW Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;  
 KW cerebral ischaemia; peripheral neuropathy; Alzheimer's disease;  
 KW Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; enzyme.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 138..152  
 FT /label= Glycoprotease\_domain  
 XX  
 WO200277233-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 08-NOV-2001; 2001WO-US046724.  
 XX  
 PR 08-NOV-2000; 2000US-0246768P-yes  
 PR 08-NOV-2000; 2000US-0246772P-yes  
 PR 15-NOV-2000; 2000US-0249185P.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 PI Leyby KR, Kapeller-Libermann R, Glucksmann M;  
 XX WPI: 2003-029938/02.  
 DR N-PSDB; ABS57020.  
 XX  
 PT New adenosine deaminase, glycoprotease and seven transmembrane domain  
 PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,  
 PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or  
 PT hypertension.  
 XX  
 PS Claim 4; Fig 8A-B; 178pp; English.  
 XX  
 CC The invention describes isolated 38650, 28472, 5495, 65507, 81588 and  
 CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The  
 CC 38650 nucleic acid molecule comprises a sequence encoding adenosine  
 CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding  
 CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise  
 CC sequences that encode a human seven transmembrane domain (7TM). The  
 CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide  
 CC sequences are useful for diagnosing, preventing or treating a subject  
 CC with or at risk of developing a disorder, e.g. cancer or aberrant  
 CC cellular proliferation and/or differentiation (e.g. breast, ovarian,  
 CC prostate, colon or lung cancer), immune disorders, heart disorders,  
 CC cardiovascular disorders, endothelial disorders, hemopoietic disorders,  
 CC blood vessel disorders, brain disorders, pain and metabolic disorders,  
 CC liver disorders or platelet disorders. These disorders include carcinoma,  
 CC sarcoma, leukaemia, Hodgkin's disease, autoimmune disorders,

CC hypertension, atherosclerosis, heart failure, myocardial infarction,  
 CC ischaemic heart disease, Crohn's disease, Grave's disease, Kawasaki  
 CC syndrome, Raynaud's disease, aneurysm, cerebral ischaemia, peripheral  
 CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,  
 CC cachexia or diabetes. This is the amino acid sequence of the novel human  
 CC glycoprotease 28472  
 XX  
 SQ Sequence 414 AA;  
 Query Match 98.3%; Score 2088; DB 6; Length 414;  
 Best Local Similarity 98.3%; Pred. No. 1.9e-211;  
 Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 MLITKTAGVFFPKSRKRYEFLRSFNFHGTLLFLHKIVLGITSCDDTAAAVVDETGVN 60  
 DB 1 MLITKTAGVFFPKSRKRYEFLRSFNFHGTLLFLHKIVLGITSCDDTAAAVVDETGVN 60  
 QY 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIQVEALSASGVSPSDLSAIAITTKPGL 120  
 DB 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIQVEALSASGVSPSDLSAIAITTKPGL 120  
 QY 121 ALSIGVLSFSLQVLGQKKPFIPIHMEAHALTIRLTNKEVPPFLVLLISGHCILLAV 180  
 DB 121 ALSIGVLSFSLQVLGQKKPFIPIHMEAHALTIRLTNKEVPPFLVLLISGHCILLAV 180  
 QY 181 QGVSDFLILGKSLDIAPGMDLVKVARSLIKHPECSSTSGGKAIEMHAKQGRFFFDIK 240  
 DB 181 QGVSDFLILGKSLDIAPGMDLVKVARSLIKHPECSSTSGGKAIEMHAKQGRFFFDIK 240  
 QY 241 PPLHAKKCDPSFTGLQHTVDTKIMKKEKEGIEKQIISADIAATVQHTMACHLVKR 300  
 DB 241 PPLHAKKCDPSFTGLQHTVDTKIMKKEKEGIEKQIISADIAATVQHTMACHLVKR 300  
 QY 301 THRALLFCQRDLTPONNAVIVASGVASNFYIRALLETITNAOCTLLCPPRLLCTDNG 360  
 DB 301 THRALLFCQRDLTPONNAVIVASGVASNFYIRALLETITNAOCTLLCPPRLLCTDNG 360  
 QY 361 IMIANNGIERLRAGILHDIGIRYEPKCPGLVDISKEVGASIKVPOLKMEI 414  
 DB 361 IMIANNGIERLRAGILHDIGIRYEPKCPGLVDISKEVGASIKVPOLKMEI 414  
 RESULT 6  
 ABU09569 standard; protein; 414 AA.  
 XX  
 AC ABU09569;  
 XX  
 DT 08-JUL-2003 (first entry)  
 XX  
 DE Human glycoprotease encoded by cDNA 28472.  
 XX  
 KW Human; enzyme; cancer; aberrant cellular proliferation; differentiation;  
 KW immune disorders; heart disorder; brain disorder;  
 KW cardiovascular disorder; endothelial cell disorder; pain disorder;  
 KW haematopoietic disorder; blood vessel disorder; metabolic disorder;  
 KW liver disorder; platelet disorder; glycoprotease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US200309017-A1.  
 XX  
 PD 09-JAN-2003.  
 XX  
 PE 08-NOV-2001; 2001US-00012140.  
 XX  
 PR 08-NOV-2000; 2000US-0246768P.  
 PR 08-NOV-2000; 2000US-0246772P. **YES**  
 PR 15-NOV-2000; 2000US-0249185P.  
 XX  
 XX (LEIBY/ LEIBY K R.  
 PA (KAPE/) KAPPELLER-LIBERMANN R.  
 PA (GLUC/) GLUCKSMANN M A.

XX  
 PT Leiby KR, Kapeller-Libermann R, Glucksmann MA;  
 XX  
 XX WPI; 2003-42888/40.  
 DR N-PSDB; ACA60887.  
 XX  
 PT New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid  
 PT molecules, useful for diagnosing, treating cancer, pain, or immune,  
 PT heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic  
 PT and liver disorders.  
 XX  
 PS Claim 4; Fig 8; 90pp; English.  
 XX  
 CC The invention relates to an isolated 38650 (encoding adenosine  
 CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7  
 CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or  
 CC a sequence which is at least 60% identical to the six nucleic acids or  
 CC their open reading frames, fragments of at least 15 nucleotides,  
 CC naturally occurring variants, or a DNA insert of the plasmid deposited  
 CC with the American Type Culture Collection as Accession No. not defined in  
 CC the specification, which encodes the amino acid sequence). Also included  
 CC are a host cell containing the nucleic acids (used to produce the  
 CC proteins), the encoded proteins, an antibody that selectively binds to  
 CC the polypeptide, and identifying a compound that binds to/modulates the  
 CC activity of the polypeptide. The nucleic acid molecules, polypeptides and  
 CC methods are useful for diagnosing, treating cancer, aberrant cellular  
 CC proliferation and/or differentiation, immune disorders, heart disorders,  
 CC cardiovascular disorders including endothelial cell disorders,  
 CC haematopoietic disorders, blood vessel disorders, brain disorders, pain  
 CC and metabolic disorders, liver disorders and platelet disorders (many  
 CC examples of these disorders are given in the specification). The present  
 CC sequence is the Human glycoprotease encoded by cDNA 28472  
 XX  
 SQ Sequence 414 AA;  
 Query Match 98.3%; Score 2088; DB 6; Length 414;  
 Best Local Similarity 98.3%; Pred. No. 1.9e-211;  
 Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 MLITKTAGVFFPKSRKRYEFLRSFNFHGTLLFLHKIVLGITSCDDTAAAVVDETGVN 60  
 DB 1 MLITKTAGVFFPKSRKRYEFLRSFNFHGTLLFLHKIVLGITSCDDTAAAVVDETGVN 60  
 QY 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIQVEALSASGVSPSDLSAIAITTKPGL 120  
 DB 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIQVEALSASGVSPSDLSAIAITTKPGL 120  
 QY 121 ALSIGVLSFSLQVLGQKKPFIPIHMEAHALTIRLTNKEVPPFLVLLISGHCILLAV 180  
 DB 121 ALSIGVLSFSLQVLGQKKPFIPIHMEAHALTIRLTNKEVPPFLVLLISGHCILLAV 180  
 QY 181 QGVSDFLILGKSLDIAPGMDLVKVARSLIKHPECSSTSGGKAIEMHAKQGRFFFDIK 240  
 DB 181 QGVSDFLILGKSLDIAPGMDLVKVARSLIKHPECSSTSGGKAIEMHAKQGRFFFDIK 240  
 QY 241 PPLHAKKCDPSFTGLQHTVDTKIMKKEKEGIEKQIISADIAATVQHTMACHLVKR 300  
 DB 241 PPLHAKKCDPSFTGLQHTVDTKIMKKEKEGIEKQIISADIAATVQHTMACHLVKR 300  
 QY 301 THRALLFCQRDLTPONNAVIVASGVASNFYIRALLETITNAOCTLLCPPRLLCTDNG 360  
 DB 301 THRALLFCQRDLTPONNAVIVASGVASNFYIRALLETITNAOCTLLCPPRLLCTDNG 360  
 QY 361 IMIANNGIERLRAGILHDIGIRYEPKCPGLVDISKEVGASIKVPOLKMEI 414  
 DB 361 IMIANNGIERLRAGILHDIGIRYEPKCPGLVDISKEVGASIKVPOLKMEI 414  
 RESULT 7  
 ADA54471 standard; protein; 364 AA.  
 ID ADA54471  
 XX  
 AC ADA54471;

XX 20-NOV-2003 (first entry)  
 XX Human protein, SEQ ID 2039.  
 DE  
 XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
 KM Gene Therapy; human; secretory protein; membrane proteins; cancer;  
 KM inflammatory disease; osteoporosis; neurological disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI293569-A2.  
 XX  
 PD 19-MAR-2003.  
 XX  
 PF 21-MAR-2002; 2002EP-00006586.  
 XX  
 PR 14-SEP-2001; 2001JP-00328381.  
 PR 24-JAN-2002; 2002US-0350435P.  
 XX  
 PA (HELI-) HELIX RES INST.  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 PI Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;  
 XX  
 DR WPI; 2003-395539/38.  
 DR N-PSDB; ADA52832.  
 XX  
 PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
 PT and/or membrane proteins, useful for developing medicines for diseases in  
 PT which the gene is involved, or as target molecules for gene therapy.  
 XX  
 PS Claim 14; SEQ ID NO 2039; 205pp; English.  
 XX  
 CC The present invention relates to novel human secretory or membrane  
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
 CC ADA54071). The coding sequences are useful in the gene therapy of  
 CC diseases caused by abnormalities of the proteins, e.g. cancer,  
 CC inflammatory diseases, osteoporosis or neurological disease.  
 CC  
 XX  
 SQ Sequence 364 AA;  
 Query Match 86.8%; Score 1845; DB 6; Length 364;  
 Best Local Similarity 99.2%; Pred. No. 7, 9e-186;  
 Matches 361; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MLITKTGVFPKPSRKRYEFLRSFNPFGTLFLHKIYLGIEETSCDDTAATAAVDETGNV 60  
 DB 1 MLITKTGVFPKPSRKRYEFLRSFNPFGTLFLHKIYLGIEETSCDDTAATAAVDETGNV 60  
 QY 61 LGEAHSOTEVHLLKTGIVPPAAQOLHRENIORIVQEAALSASGVSPDSALATTTKPGI 120  
 DB 61 LGEAHSOTEVHLLKTGIVPPAAQOLHRENIORIVQEAALSASGVSPDSALATTTKPGI 120  
 QY 121 ALSIGVGSFSLQVLQGLKKPPIPIHMEAAHLLTRLNKVEFPPIVLLISGCHLLALV 180  
 DB 121 ALSIGVGSFSLQVLQGLKKPPIPIHMEAAHLLTRLNKVEFPPIVLLISGCHLLALV 180  
 QY 121 ALSLGVGSFSLQVLQGLKKPPIPIHMEAAHLLTRLNKVEFPPIVLLISGCHLLALV 180  
 DB 121 ALSLGVGSFSLQVLQGLKKPPIPIHMEAAHLLTRLNKVEFPPIVLLISGCHLLALV 180  
 QY 181 QGVSPFLLIGKSLDIPGMILKVAARLPLIKHPECTWSGGALSHLAKQGRFPHDIK 240  
 DB 181 QGVSPFLLIGKSLDIPGMILKVAARLPLIKHPECTWSGGALSHLAKQGRFPHDIK 240  
 QY 241 PPIHAKKNCDFSTGLOHTYTDKIIMKEKEBEGIEKQIISADIAATVOHTACHLVKR 300  
 DB 241 PPIHAKKNCDFSTGLOHTYTDKIIMKEKEBEGIEKQIISADIAATVOHTACHLVKR 300  
 QY 301 THRAILFCQKORDLPPONNAVLVAGGVASNFYIRRALILTNATQCTLLCPPRLCTDNG 360  
 DB 301 THRAILFCQKORDLPPONNAVLVAGGVASNFYIRRALILTNATQCTLLCPPRLCTDNG 360  
 QY 361 IMIA 364

DB 361 IMIA 364  
 RESULT 8  
 ABG96487  
 ID ABG96487 standard; protein; 267 AA.  
 XX  
 AC ABG96487;  
 XX  
 DT 11-DEC-2002 (first entry)  
 XX  
 DE Novel human metalloprotease MPI fragment #1.  
 XX  
 KM Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
 KM motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
 KM reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
 KM genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
 KM Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
 KM Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
 KM liver disease; renal disease; immune disorder; rheumatoid arthritis;  
 KM acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
 KM emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
 KM neurological disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200272751-A2.  
 XX  
 PD 19-SEP-2002.  
 XX  
 PF 05-FEB-2002; 2002WO-US003353.  
 XX  
 PR 05-FEB-2001; 2001US-0266518P.  
 PR 10-APR-2001; 2001US-0282814P.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;  
 PI  
 DR WPI; 2002-723329/78.  
 DR N-PSDB; ABS76639.  
 XX  
 PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
 PT treating, or ameliorating diseases associated with aberrant  
 PT metalloprotease activity, e.g. immune, metabolic, inflammatory and  
 PT neurological disorders.  
 XX  
 PS Claim 5; Page 29; 473pp; English.  
 XX  
 CC The invention describes an isolated nucleic acid molecule (I) encoding a  
 CC metalloprotease (MP-1). (I) is useful for preventing, treating, or  
 CC ameliorating a medical condition, particularly an immune disorder, an  
 CC aberrant glutamate transport or motor neuron disorder, such as  
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
 CC condition. The compositions and methods are also useful for diagnosing,  
 CC prognosticating, treating, ameliorating and/or treating disorders  
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
 CC or Huntington's disease), neurodegenerative diseases (Alzheimer's disease,  
 CC or Huntington's disease), Huntington's disease or Tourette syndrome, liver  
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
 CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
 CC sequence of a metalloprotease MPI protein  
 XX  
 SQ Sequence 267 AA;  
 Query Match 65.2%; Score 1385; DB 5; Length 267;  
 Best Local Similarity 100.0%; Pred. No. 2, 4e-137;  
 Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;











PR	15-JUL-1999	99SUS-0144005P
PR	16-JUL-1999	99SUS-0144008P
PR	16-JUL-1999	99SUS-0144086P
PR	19-JUL-1999	99SUS-0144325P
PR	19-JUL-1999	99SUS-0144331P
PR	19-JUL-1999	99SUS-0144332P
PR	19-JUL-1999	99SUS-0144333P
PR	19-JUL-1999	99SUS-0144334P
PR	19-JUL-1999	99SUS-0144335P
PR	20-JUL-1999	99SUS-0144352P
PR	20-JUL-1999	99SUS-0144384P
PR	20-JUL-1999	99SUS-0144418P
PR	21-JUL-1999	99SUS-0145086P
PR	21-JUL-1999	99SUS-0145088P
PR	21-JUL-1999	99SUS-0145089P
PR	22-JUL-1999	99SUS-0145087P
PR	22-JUL-1999	99SUS-0145089P
PR	22-JUL-1999	99SUS-0145192P
PR	23-JUL-1999	99SUS-0145145P
PR	23-JUL-1999	99SUS-0145188P
PR	23-JUL-1999	99SUS-0145242P
PR	26-JUL-1999	99SUS-0145276P
PR	27-JUL-1999	99SUS-0145913P
PR	27-JUL-1999	99SUS-0145918P
PR	27-JUL-1999	99SUS-0145919P
PR	28-JUL-1999	99SUS-0145581P
PR	02-AUG-1999	99SUS-0146386P
PR	02-AUG-1999	99SUS-0146388P
PR	02-AUG-1999	99SUS-0146389P
PR	03-AUG-1999	99SUS-0147038P
PR	04-AUG-1999	99SUS-0147204P

PR	06-Oct-1999	99US-0157865P
PR	07-Oct-1999	99US-0158029P
PR	08-Oct-1999	99US-0158532P
PR	12-Oct-1999	99US-0158369P
PR	13-Oct-1999	99US-0158923P
PR	13-Oct-1999	99US-0159294P
PR	13-Oct-1999	99US-0159295P
PR	14-Oct-1999	99US-0159329P
PR	14-Oct-1999	99US-0159330P
PR	14-Oct-1999	99US-0159331P
PR	14-Oct-1999	99US-0159637P
PR	14-Oct-1999	99US-0159638P
PR	14-Oct-1999	99US-0159584P
PR	18-Oct-1999	99US-0160741P
PR	21-Oct-1999	99US-0160767P
PR	21-Oct-1999	99US-0160768P
PR	21-Oct-1999	99US-0160770P
PR	21-Oct-1999	99US-0160814P
PR	21-Oct-1999	99US-0160815P
PR	22-Oct-1999	99US-0160980P
PR	22-Oct-1999	99US-0160981P
PR	22-Oct-1999	99US-0161049P
PR	22-Oct-1999	99US-0161049P
PR	25-Oct-1999	99US-0161405P
PR	25-Oct-1999	99US-0161406P
PR	26-Oct-1999	99US-0161359P
PR	26-Oct-1999	99US-0161360P
PR	26-Oct-1999	99US-0161361P
PR	28-Oct-1999	99US-0161920P
PR	28-Oct-1999	99US-0161922P
PR	28-Oct-1999	99US-0161933P
PR	29-Oct-1999	99US-0162142P

Query Match 31.0%; Score 659.5; DB 3; Length 444;

Best Local Similarity 40.8%; Pred. No. 3.3e-60;  
Matches 148; Conservative 57; Mismatches 127; Indels 31; Gaps 6.

38 IVLGIETSCDPTAAAVVDETGNVLGEAIIHSCTEVHLKTGGIVPPAAOOLHRENIORIVOE 97

PR	09-AUG-1999	99US-0147493P
PR	09-AUG-1999	99US-0147935P
PR	10-AUG-1999	99US-0148171P
PR	11-AUG-1999	99US-0148319P
PR	12-AUG-1999	99US-0148341P
PR	13-AUG-1999	99US-0148655P
PR	13-AUG-1999	99US-0148684P
PR	16-AUG-1999	99US-0149368P
PR	17-AUG-1999	99US-0149175P
PR	18-AUG-1999	99US-0149426P
PR	20-AUG-1999	99US-0149722P
PR	20-AUG-1999	99US-0149723P
PR	20-AUG-1999	99US-0149929P
PR	23-AUG-1999	99US-0149902P
PR	23-AUG-1999	99US-0149930P
PR	25-AUG-1999	99US-0150566P
PR	26-AUG-1999	99US-0150884P
PR	27-AUG-1999	99US-0151065P
PR	27-AUG-1999	99US-0151066P
PR	27-AUG-1999	99US-0151080P
PR	30-AUG-1999	99US-0151303P
PR	31-AUG-1999	99US-0151438P
PR	01-SEP-1999	99US-0151930P
PR	01-SEP-1999	99US-0152363P
PR	07-SEP-1999	99US-0153070P
PR	10-SEP-1999	99US-0153758P
PR	15-SEP-1999	99US-0154018P
PR	16-SEP-1999	99US-0154039P
PR	20-SEP-1999	99US-0154707P
PR	22-SEP-1999	99US-0155139P
PR	23-SEP-1999	99US-0155486P
PR	24-SEP-1999	99US-0155659P
PR	26-SEP-1999	99US-0156458P
PR	28-SEP-1999	99US-0156566P
PR	05-OCT-1999	99US-0157113P
PR	05-OCT-1999	99US-0157753P

RESULT	12
AAV52216	
ID	AAV52216 standard; protein; 463 AA
XX	
AC	
XX	AAV52216;
XX	
DT	09-FEB-2000 (first entry)

QY 277 QILSSPADIAATVQHTMACHLVKTRTRAILFCKQRDLDLPONNAVLVAGGVASNFYIRRA 336  
 Db 309 --IRNRADIASFOEVAVLHEEKCERAIIDMALE---LEPSIKHWIISGGVANSKYRLR 363  
 QY 337 LEILINAOQCTLLCPRPRLCTONGMIANGIERLRAGIGILHDIGIYE----- 387  
 Db 364 LNNIVENKNLKVAVCPSPSLCTDNGVAVMTGLEHFRVG-----RYDPPPEATEPE 413  
 QY 388 -----PKCPPLGVDISKEGEVA 403  
 Db 414 DYVYDLRRPRWPLGEEYAKGRSEA 436  
 RESULT 13  
 ABG96489  
 ID ABG96489 standard, protein, 179 AA.  
 AC ABG96489;  
 XX  
 DT 11-DEC-2002 (first entry)  
 XX  
 DE Novel human metalloprotease MPI fragment #3.  
 XX  
 KM Metalloprotease, MP-1; immune disorder; glutamate transport; cancer;  
 KM motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
 KM reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
 KM genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
 KM Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
 KM Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
 KM liver disease; renal disease; immune disorder; Rheumatoid arthritis;  
 KM acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
 KM emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
 XX neurological disorder.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200272751-A2.  
 XX  
 PD 19-SEP-2002.  
 XX  
 PF 05-FEB-2002; 2002WO-US003353.  
 XX  
 PR 05-FEB-2001; 2001US-0266518P.  
 PR 10-APR-2001; 2001US-0282814P.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S,  
 DR MPI; 2002-723329/78.  
 XX  
 PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
 PT treating, or ameliorating diseases associated with aberrant  
 PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and  
 PT neurological disorders.  
 XX  
 PS Disclosure; Page 50; 473pp; English.

CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
CC sequence of a metalloprotease MPl protein  
XX  
SQ Sequence 179 AA;

Query Match 25.8%; Score 549; DB 5; Length 179;  
Best Local Similarity 66.5%; Pred. No. 3.8e-49;  
Matches 119; Conservative 0; Mismatches 0; Indels 60; Gaps 1;

QY 38 IVLGIEISCDPTAAVDETVNVLGEAHSOTFVHLKTKGIVPPAAQQLHRENIQVGE 97  
DB 1 IVLGIEISCDPTAAVDETVNVLGEAHSOTFVHLKTKGIVPPAAQQLHRENIQVGE 60  
QY 98 ALSASGVSPDLSAIAITTIKPGIALSLGVLSFSLQVLQVQLKKPFI 144  
DB 61 ALSASGVSPDLSAIAITTIKPGIALSLGVLSFSLQVLQVQLKKPFI 120  
QY 145 -----IHMEAHALTIR 156  
DB 121 GGAGGCTCATGCACTTACTATTAGGTGACCAATAAAGTAGAATTTCIHMEAHALTIR 179

RESULT 14  
ABU22934  
ID ABU22934 standard; protein; 382 AA.

AC ABU22934;

DT 19-JUN-2003 (first entry)

DE Protein encoded by prokaryotic essential gene #8461.

KM Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Bordetella pertussis.

PN WO20027183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

PA (ELITRA) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyekind JW,

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

PI MPI; 2003-029926/02.

PI N-PSDB; ACA26804.

PI New antisense nucleic acids, useful for identifying proteins or screening

PI for homologous nucleic acids required for cellular proliferation to

PI isolate candidate molecules for rational drug discovery programs.

PI Claim 25; SEQ ID NO 50858; 1766gp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation and the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (6)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIP0 at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 382 AA;

Query Match 24.7%; Score 524; DB 6; Length 382;  
Best Local Similarity 37.3%; Pred. No. 5.7e-46;  
Matches 134; Conservative 63; Mismatches 136; Indels 26; Gaps 10;

QY 29 HPG--TLPLHK---IVLGIEISCDPTAAVDETVNVLGEAHSOTFVHLKTKGIVPPAA 83

DB 24 HPGPRTLVNVSAPMITLGFESSCDPTGVAVCTERGLAHALMTQIAMQBYGVGVBELA 83

QY 84 QQLHRENIQVGEALSASGVSPDLSAIAITTIKPGIALSLGVLSFSLQVLQVQLKKPFI 143

DB 84 SRDHIRRVPLTRQVLAENGLTLADGAVATVAGPGLAGLTVGASVAAQLAWSRBLPAI 143

QY 144 PIHMEAHALTIRLTN-KVEPPVLIISSGHCLLAVQGVSPFLIGKSLDIAPEGMD 202

DB 144 GIHMLEGHLISPLAEPREPFVALLVSGHQOLMVDGVGYELLGETLIDPAAGBAFD 203

QY 203 KVARRLSLIKHPECSTMSGKAEHLAKQGNRTHFPIKPELHNAKDSFSTGLQ-HVTD 261

DB 204 KSAKLMGL-GYP-----GGPALARLAEQDASRYDLPRLHLSGDSFSGIKTAVLT 256

QY 262 KITMKKEKEGIEKQOILSSAADIATVQHTMACHLVKTRTHAILFCKQRDLFPQNNAVL 321

DB 257 RV--KAATRDGSLGR--QDRADLAATQAIVYLAAKIRAL---KQGL-----RL 304

QY 322 VASGVASNFTYIRALEILTNAQTLLCPPRLCTDNGIMAMGIERLPAGIGIILD 380

DB 305 VVAGVGVANALLRAHLARALKPRAEAYFPPLSLCTDNGAMIAFAAERYVAGIADURE 363

RESULT 15

ABG96491  
ID ABG96491 standard; protein; 421 AA.

AC ABG96491;

DT 11-DEC-2002 (first entry)

DE Novel human metalloprotease associated protein #2.

KM Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;

KM motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;

KM reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;

KM genital wart; metabolic disorder; premature puberty; Kallman syndrome;

KM Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;

KM Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;

KM liver disease; renal disease; immune disorder; rheumatoid arthritis;

KM acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;

KM emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;

XX neurological disorder.





Db 414 DYVYDLRPRWPLGEEYAKGRSEA 436

## RESULT 2

E71711 Probable O-sialoglycoprotein endopeptidase (gcp) RP037 - Rickettsia prowazekii

C/Species: Rickettsia prowazekii

C/Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 09-Jul-2004

C/Accession: E71711

R/Andersson, S.G.E.; Zomorodipour, A.; Andersson, U.O.; Sichteritz-Ponten, T.; Alsmark, U.

Nature 396, 133-140, 1998

A/Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A/Reference number: A71630; MUID:99039499; PMID:9823893

A/Accession: E71711

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-387 <AND>

A/Cross-references: UNIPROT:Q9ZEAB; GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CA1450

A/Experimental source: strain Madrid E

C/Genetics:

A/Gene: gcp; RP037

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 26.0%; Score 553; DB 2; Length 387;

Best Local Similarity 35.1%; Pred. No. 5,6e-37;

Matches 133; Conservative 56; Mismatches 130; Indels 60; Gaps 6;

QY 37 KIVVIGTSCDDTAAAVVDEGTVLGEAHSQTEVHLKTCGIVPPAAQOLHRENIQRIYQ 96

Db 2 KKLIGISSCDDTASITERRKILSNITISQTEHAVFGVGPPIAARSHLSNDQAL 61

QY 97 EALSASGVSPSDLSAIAITTKPGIALSLGVGLSFLQVGLKPPFPIHMEAHALFTR 156

Db 62 NTLKKSMTLETFEISALITSGPGLIGVIVSGMPKRSLSALKKPFIAINHEGHALTR 121

QY 157 LTNKVEPFLVLLISGGHCLALVQVSDPFLGKSLDIAPGMDLVKARLSLTKHPEC 216

Db 122 LFDNISVYLLLLASGGHCGQFVAVLGKVKILGTIDAVGETDKVAKMLNLT----- 175

QY 217 STMSGKALIEHLAKQGNRFHFDIKPPLHAKNCDFSTGLQHTVDKIIMK-KEKEBGIEK 275

Db 176 -SFGSGEIEERAKLGNPHKXKPPKPIINSQCNMSFGLKTAIVRLLIMNLKEVNDVSV- 232

QY 276 GQLISSADIAATVQHTMACLVKTRHALLFCQ-----RD-- 312

Db 233 -----INDIASFOFTGAILSSKQDAIRLYKQIILNDYEDINHPTKLNLSFRKDEF 286

QY 313 -----LLPON-----NAVLVSGVANSFYIRALAILTNATQCTLLCP 351

Db 287 NMRKLECTPRKRIHIONSYRSNLNLTIVAGVANKYQELISDCTREYGRLLAP 346

QY 352 PPRICTDNGIMIANNGIER 370

Db 347 PMHLCTDMAAMIAVAGLER 365

## RESULT 3

AB2902 O-sialoglycoprotein endopeptidase gcp [imported] - Agrobacterium tumefaciens (strain C58

C/Species: Agrobacterium tumefaciens

C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002

C/Accession: AB2902

R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavyn, T.; Levy, R.; Li, M.; McClell

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A/Reference number: AB2977; MUID:21608550; PMID:11743193

A/Accession: AB2902

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-365 <KUR>

A/Cross-references: GB:AE008688; PIDN:AA143632.1; PID:g17741154; GSPDB:GN00186

A/Experimental source: strain C58 (Dupont)

C/Genetics:

A/Gene: gcp

A/Map position: circular chromosome

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 25.8%; Score 548; DB 2; Length 365;

Best Local Similarity 36.9%; Pred. No. 1.3e-36;

Matches 137; Conservative 58; Mismatches 142; Indels 34; Gaps 8;

QY 39 VIGIETSCDDTAAAVV---DEGTVLGEAHSQTEVHLKTCGIVPPAAQOLHRENIQRI 94

Db 7 ILGIFTSCEETASIVVRADGRGEIVSDVLSQLEHSAGVGVPPIAARAHVALDLT 66

QY 95 VQELASASGVSPSDLSAIAITTKPGIALSLGVGLSFLQVGLKPPFPIHMEAHALF 154

Db 67 VEEALDQAGVKLADVDALAAATSGPGLIGLVLGMLGKAKAAKRLVAINHLEGHALT 126

QY 155 IRLTNKVEPFLVLLISGGHCLALVQVSDPFLGKSLDIAPGMDLVKARLSLTKHP 214

Db 127 ARLTDLSPFYLMLLVSGHGTQLVLRGVGERYERWGTITDDALGEAFDITAKLGL-PIV 185

QY 215 ECSTMSGKALIEHLAKQGNRFHFDIKPPLHAKNCDFSTGL---QHTVDKIIMKKE 270

Db 186 -----GCPAVENMAAKGDPDRFPPLRPVWGEARLDFSSGLKTAVRQAATAIAPLSEQD 239

QY 271 EGIEKQILISSADIAATVQHTMACLVKTRHALLFCQORDLLPONNA--VLVAGGVA 328

Db 240 -----IADICASFQKAVSRTLDKRIGRGLARFVE--FPHINGPALVAVGVA 286

QY 329 SNFYIRALEITLMAATQCTLLCPPRICTDNGIMIANNGIERLRAIGILHDIETRIYEP 388

Db 287 ANQELRQTLQALCDTHGFRFVAPPHRLCTDMAAMIAWAGERMAEG---RQDALVELAP 342

QY 389 KCPLGVDISKE 399

Db 343 RSRWPLDGSAE 353

## RESULT 4

D97677 Probable O-sialoglycoprotein endopeptidase (glycoproteinase) [imported] - Agrobacterium

C/Species: Agrobacterium tumefaciens

C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002

C/Accession: D97677

R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirio, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lapas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A/Reference number: A97359; MUID:21608551; PMID:11743194

A/Accession: D97677

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-366 <KUR>

A/Cross-references: GB:AE007869; PIDN:AAK88373.1; PID:g15157858; GSPDB:GN00169

C/Genetics:

A/Gene: AGR\_C 4806

A/Map position: circular chromosome

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 25.8%; Score 548; DB 2; Length 366;

Best Local Similarity 36.9%; Pred. No. 1.3e-36;

Matches 137; Conservative 58; Mismatches 142; Indels 34; Gaps 8;

QY 39 VIGIETSCDDTAAAVV---DEGTVLGEAHSQTEVHLKTCGIVPPAAQOLHRENIQRI 94

Db 8 ILGIFTSCEETASIVVRADGRGEIVSDVLSQLEHSAGVGVPPIAARAHVALDLT 67

QY 95 VQELASASGVSPSDLSAIAITTKPGIALSLGVGLSFLQVGLKPPFPIHMEAHALF 154

Db 68 VEEALDQAGVKLADVDALAAATSGPGLIGLVLGMLGKAKAAKRLVAINHLEGHALT 127







Matches	136;	Conservative	54;	Mismatches	144;	Indels	25;	Gaps	7;		
QY	38	IVLGIEHS	CDPTAAAV----	DEGNVL	LGAEI	HSQTEVHL	KTGGIV	PPAAQOL	RENIORI	93	
				:	:	:	:	:	:		
Db	10	LILGLEFS	CDCTEASV	VRRAAD	GTVTYLS	VTGTFQ	FEHNA	PFGGV	VEI	14	
				:	:	:	:	:	:		
QY	94	IVQGLAS	GVSP	SDLSA	ITTTIK	PGIALSL	GVGS	FSFLQ	VGQKK	PFPIH	153
				:	:	:	:	:	:		
Db	70	IAAEV	APAAV	GFED	LDG	VAAT	AGPGL	VGV	MGAL	RGKAL	129
				:	:	:	:	:	:		
QY	154	TIRLTK	VEP	FPVL	VLIS	GGH	CLL	LVQ	VSDF	LLG	213
				:	:	:	:	:	:		
Db	130	SARLG	ADIA	VFP	LLV	SGCH	QCL	LEV	SGV	ACR	188
				:	:	:	:	:	:		
QY	214	PEG	CTMG	GKAI	IEHL	AKGN	FHF	DIKE	PL	HLAK	273
				:	:	:	:	:	:		
Db	189	P-----	GGP	LEK	LA	VQGD	PT	RYAL	P	AL	240
				:	:	:	:	:	:		
QY	274	EKG	IL	SSA	DI	AA	ITV	OH	TMA	CH	333
				:	:	:	:	:	:		
Db	241	-----	ARR	DLA	GV	QAL	AR	Q	L	R	291
				:	:	:	:	:	:		
QY	334	RA	LEI	IT	NA	O	CT	LL	C	P	392
				:	:	:	:	:	:		
Db	292	RA	ALL	AD	CE	K	KG	F	P	A	347
				:	:	:	:	:	:		
<p>RESULT 8</p> <p>T18825  hypothetical protein C01G10.10 - Caenorhabditis elegans  C:Species: Caenorhabditis elegans  C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000  C:Accession: T18825  R:Matthews, L.  submitted to the EMBL Data Library, October 1996  A:Reference number: Z19027  A:Accession: T18825  A:Status: preliminary; translated from GB/EMBL/DBJ  A:Molecule type: DNA  A:Residues: 1,421 &lt;MBL&gt;  A:Cross-references: EMBL:Z81030; PIDN:CA802716.1; GSPDB:GN00023; CESP:C01G10.10  A:Experimental source: clone C01G10  C:Genetics:  A:Gene: CESP:C01G10.10  A:Map position: 5  A:Intons: 31/2; 72/3; 122/2; 177/1; 272/3; 315/1; 353/2  C:Superfamily: O-sialoglycoprotein endopeptidase</p>											
Query Match	24.7%	Score	524;	DB	2;	Length	421;				
Best Local Similarity	33.4%	Pred. NO.	1.4e-34;								
Matches	130;	Conservative	77;	Mismatches	146;	Indels	36;	Gaps	8;		
QY	39	VLG	IETSC	DDT	PA	AAV	V	DE	TGN	VL	98
				:	:	:	:	:	:		
Db	25	VL	G	I	E	T	S	C	D	D	84
				:	:	:	:	:	:		
QY	99	LS	AS	G	S	P	S	D	I	S	158
				:	:	:	:	:	:		
Db	85	L	N	D	A	G	T	S	K	P	144
				:	:	:	:	:	:		
QY	159	N	K	V	E	P	P	L	V	L	217
				:	:	:	:	:	:		
Db	145	D	D	S	V	R	F	P	S	A	203
				:	:	:	:	:	:		
QY	218	T	M	S	G	K	A	I	E	H	274
				:	:	:	:	:	:		
Db	204	G	I	N	H	G	A	V	E	I	263
				:	:	:	:	:	:		
QY	275	K	G	O	L	I	S	A	D	I	334
				:	:	:	:	:	:		
Db	264	-----	I	P	D	C	A	S	I	O	313
				:	:	:	:	:	:		
QY	335	R	A	L	E	I	T	N	A	O	388
				:	:	:	:	:	:		

Db 314 GAISLTLAAHNTTTLKVLLSLCTDNAMETAYSGD-----LMLVNRSEALVWRPNDIPDT 367

Qy 389 -----KCPGLVDISKVEGEASIKVPOQKM 412

Db 368 IYAHARS DIGTDASSSEI-----IDTPRRKL 392

RESULT 9

G70369  
sialoglycoproteinase - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004  
A:Accession: G70369  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov  
V.

Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: G70369  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-335 <AOF>  
A:Cross-references: UNIPROT:O66986; GB:AE00708; NID:g2983356; PIDN:AAC06951.1; PID:g298  
A:Experimental source: strain VFS  
C:Genetics:  
A:Gene: gcp  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 24.3%; Score 516.5; DB 2; Length 335;  
Best Local Similarity 36.5%; Pred. No. 4,2e-34;  
Matches 130; Conservative 66; Mismatches 125; Indels 35; Gaps 10;

Qy 40 LGIETSCDDTAADVDETNVIGEAHISQTEVHLKTGGIYPPAAOQLHRENIQRIYQEA 99

Db 4 LAVETSCDETAIAIYDDQKGVIGNVLLSCAVVHSPFGVVPDELARERTRNILPIFDRL 63

Qy 100 SASGVSPPDLASATIRIKGLALSLGVGSFSFLQVQQLKKPFIPIHMEAHALITRLTN 159

Db 64 KESRINLEBIDITSPFLTPGLIISLVGVAFAPAKLAVERYPLVPHNLBGHIYSVFLEK 123

Qy 160 KVEPFLVLLISGGHCLALVQGVDFLLLGKSLIDIAFGMDLKVARRSLIKHPECSTM 219

Db 124 KYEYFFLLIISGGHLDYLVPRFGRYDFLGCTIDDAVGEAYDKAAKMLGL-GYF----- 177

Qy 220 SGGAIEHLAKQGNRFHFDIKPPLHAKNCDFSTGLQHTVTDKIIIMKKEBGEIEKGQTL 279

Db 178 -GGPIIDRLAKGGKRL-YPLPKPLMERGNINFSFGLK---TALINLLKKEGNVRK---- 228

Qy 280 SSAADIAALVQHTMACHLVKTRTRALLFCQKQDRLIPQNNAVLVASGVASNYIRRALEI 339

Db 229 --EIIAYSFQETVAVILERS---LWAKKGTGIKR---LVVGVGSANSRLR--EV 274

Qy 340 LTNATQ---CTLLCPPEPLCTDNGIMIAWNGIEBIRAGILGILHDEIGRIYEBKCP 392

Db 275 FKKAQOEYFELYIPIHPSLSTDNALMIAYAGMEFKGVAPLDVNP---QNIPL 327

RESULT 10

H83572  
O-sialoglycoprotein endopeptidase PA0580 [imported] - Pseudomonas aeruginosa (strain PAO  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
A:Accession: H83572  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bra  
dman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Larbig, K.; Lim,  
J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: H83572  
A:Status: preliminary  
A:Molecule type: DNA



A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: F65094  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Title: Preliminary: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-337 <BLAT>  
A:Cross-references: UNIPROT:P05852; GB:AE000388; GB:U00096; NID:g1789441; PIDN:AACT6100.  
A:Experimental source: strain K-12, substrain MG1655  
R:Neim, M.; Lupski, J.R.; Svec, P.; Godson, G.N.  
Gene 51, 149-161, 1987  
A:Title: Possible new genes as revealed by molecular analysis of a 5-kb Escherichia coli  
A:Reference number: A91573; MUID:87248073; PMID:3297921  
A:Accession: D29049  
A:Molecule type: DNA  
A:Residues: 1-135; 'C', 137-337 <NES>  
A:Cross-references: GB:M16194; GB:X00773; NID:g147764; PIDN:AAA72575.1; PID:G551834  
C:Genetics:  
A:Gene: yjgD  
A:Map position: 67 min  
C:Superfamily: O-sialoglycoprotein endopeptidase  
C:Keywords: hydrolase; metalloproteinase; zinc  
F:111,115/Binding site: zinc (His) #status predicted

Query Match 22.8%; Score 484.5; DB 1; Length 337;  
Best Local Similarity 35.1%; Pred. No. 1.6e-31;  
Matches 125; Conservative 54; Mismatches 148; Indels 29; Gaps 9;

Qy 39 VLGIEISCDPTAAAVDESGNVLGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEA 98  
Db 3 VLGIEISCDPTAAAVDESGNVLGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEA 62

Qy 99 LLSAGVSPSDLSAIAITTIKPGIALSLGVLSFSLQVLGQKKPFIPIHMEAHALTIRL- 157  
Db 63 LKESGLTAKDIDAVATYTAGPGLVGLALVGTGRSLAFAMVDPAIVHMEHGLAPMLE 122

Qy 158 TNKVEFPFLVLLISGGCHLALVQVSDPFLIKSLDIAFGMDLKVARRLSLIKPECS 217  
Db 123 DNPEFPFVALLVSGCHTQLISVTGIGQYELLGESIDDAAGAFDKTAKLGL-DYP--- 178

Qy 218 TMSGKAIETHLAKQGNRHFHDIKPRLHNAKCDPSTGLOHTDKIIMKKEBEGIEKQ 277  
Db 179 ---GGPMLSKMASQGTAGRFVFPFPRPMTDRPGIDFSFGSKTFAANTIRSGNDE--- 229

Qy 278 ILSSADIAATVQHTMAACHLVKTRHAILFCQKORDLIPQNN-AVLVAGGVASNFYIRRA 336  
Db 230 --QTRADIAAFEDAVDTLMTKCKRA-----LDQGRKRLVMAGGVANRTLRAX 278

Qy 337 LEILTNATQCTLLCPPLCTDNGIMIAMNGIERLRAGLGIHDIIEGIRYEPKCP 392  
Db 279 LAEMMKRRGEVFAVRFECTDNGAMIAVAGMVRFXA--GATVDL-GVSVRBRWPL 331

RESULT 14  
AG0892  
probable glycoproteinase [imported] - Salmonella enterica subsp. enterica serovar Typhi (S  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: A60892  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Comercon, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moulé, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AG0892  
A:Molecule type: DNA  
A:Status: preliminary  
A:Residues: 1-337 <PAP>  
A:Cross-references: GB:AL513382; PIDN:CAD07733.1; PID:g16504285; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY3387  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 22.8%; Score 484.5; DB 2; Length 337;  
Best Local Similarity 34.6%; Pred. No. 1.6e-31;  
Matches 123; Conservative 55; Mismatches 150; Indels 27; Gaps 8;

Qy 39 VLGIEISCDPTAAAVDESGNVLGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEA 98  
Db 3 VLGIEISCDPTAAAVDESGNVLGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEA 62

Qy 99 LLSAGVSPSDLSAIAITTIKPGIALSLGVLSFSLQVLGQKKPFIPIHMEAHALTIRL- 157  
Db 63 LKESGLTAKDIDAVATYTAGPGLVGLALVGTGRSLAFAMVDPAIVHMEHGLAPMLE 122

Qy 158 TNKVEFPFLVLLISGGCHLALVQVSDPFLIKSLDIAFGMDLKVARRLSLIKPECS 217  
Db 123 DNPEFPFVALLVSGCHTQLISVTGIGQYELLGESIDDAAGAFDKTAKLGL-DYP--- 178

Qy 218 TMSGKAIETHLAKQGNRHFHDIKPRLHNAKCDPSTGLOHTDKIIMKKEBEGIEKQ 277  
Db 179 ---GGPMLSKMASQGTAGRFVFPFPRPMTDRPGIDFSFGSKTFAANTIRSGNDE--- 229

Qy 278 ILSSADIAATVQHTMAACHLVKTRHAILFCQKORDLIPQNN-AVLVAGGVASNFYIRRA 337  
Db 230 --QTRADIAAFEDAVDTLMTKCKRA-----LDQGRKRLVMAGGVANRTLRAX 279

Qy 337 LEILTNATQCTLLCPPLCTDNGIMIAMNGIERLRAGLGIHDIIEGIRYEPKCP 392  
Db 280 LAEMMKRRGEVFAVRFECTDNGAMIAVAGMVRFXA--GATVDL-GVSVRBRWPL 331

RESULT 15  
AI0079  
probable glycoproteinase gcp [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titchell, R.W.; Holden, M.T.G.; Prentice, M.B.  
demo-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AI0079  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-337 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CA089500.1; PID:g15978736; GSPDB:GN00175  
C:Genetics:  
A:Gene: gcp  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 22.8%; Score 483.5; DB 2; Length 337;  
Best Local Similarity 34.6%; Pred. No. 2e-31;  
Matches 123; Conservative 54; Mismatches 150; Indels 29; Gaps 7;

Qy 39 VLGIEISCDPTAAAVDESGNVLGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEA 98  
Db 3 VLGIEISCDPTAAAVDESGNVLGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEA 62

Qy 99 LLSAGVSPSDLSAIAITTIKPGIALSLGVLSFSLQVLGQKKPFIPIHMEAHALTIRL- 157  
Db 63 LKESGLTAKDIDAVATYTAGPGLVGLALVGTGRSLAFAMVDPAIVHMEHGLAPMLE 122

Qy 158 TNKVEFPFLVLLISGGCHLALVQVSDPFLIKSLDIAFGMDLKVARRLSLIKPECS 217  
Db 123 DNPEFPFVALLVSGCHTQLISVTGIGQYELLGESIDDAAGAFDKTAKLGL-DYP--- 178

Qy 218 TMSGKAIETHLAKQGNRHFHDIKPRLHNAKCDPSTGLOHTDKIIMKKEBEGIEKQ 277  
Db 179 ---GGPMLSKMASQGTAGRFVFPFPRPMTDRPGIDFSFGSKTFAANTIRSGNDE--- 229

Qy 278 ILSSADIAATVQHTMAACHLVKTRHAILFCQKORDLIPQNN-AVLVAGGVASNFYIRRA 337  
Db 230 --QTRADIAAFEDAVDTLMTKCKRA-----LDQGRKRLVMAGGVANRTLRAX 279

Qy 337 LEILTNATQCTLLCPPLCTDNGIMIAMNGIERLRAGLGIHDIIEGIRYEPKCP 392  
Db 279 LAEMMKRRGEVFAVRFECTDNGAMIAVAGMVRFXA--GATVDL-GVSVRBRWPL 331

Db 230 --QTRADIARAFEDAVVDTLAIKSKRA-----LDOTGPKRLVIAGVGSANOTLRUK 278  
QY 337 LEIITNATQCTLLCPPRLCTDNGIMIAMNGIERLRAGIGIILHDI EGIRYEPKCP 392  
Db 279 LADMMOKRGEVFFYARPEFCTDNGAMIAVAGWVRLRSN---LNSELSVSVRPWPPL 331

Search completed: November 10, 2005, 16:53:07  
Job time : 26.65 secs

## **Protein Sequence Searches - February 2005**

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension .rup) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

**When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: November 10, 2005, 15:16:01 ; Search time 100.35 Seconds

(without alignments)  
2112.614 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125

Sequence: 1 MLILTKTAGVFFKPKRKRVY.....DISKEVGASIKVQLKMEI 414

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2125	100.0	414	2	096EV9
2	2090.5	98.4	439	2	09H4B0
3	1845	86.8	364	2	096NM5
4	1835	86.4	414	2	06PEB4
5	1827	86.0	414	2	08BLB6
6	1819	85.6	414	2	09D0N0
7	1725	81.2	467	2	06AYN7
8	1283	60.4	404	2	08JFW3
9	1279	60.2	404	2	08JFR7
10	772.5	36.4	401	2	07QJ98
11	714.5	33.6	409	2	09VWD6
12	668.5	31.5	480	2	022145
13	577.5	27.2	335	2	073H71
14	568.5	26.8	360	2	092LH8
15	564	26.5	362	2	098ER6
16	558	26.3	387	2	068XK3
17	553	26.0	387	1	GCP_RICPR
18	548	25.8	365	2	08UC47
19	548	25.8	366	2	07CWJ8
20	541.5	25.5	344	2	07PAG7
21	540.5	25.4	344	2	092UK6
22	539.5	25.4	359	2	08YUB1
23	539.5	25.4	359	2	08FYI5
24	538.5	25.4	323	2	0960S6
25	535.5	25.2	367	2	09ABZ9
26	531.5	25.0	389	2	06NDS4
27	524	24.7	421	2	093170
28	520	24.5	340	2	G6FCX3
29	516.5	24.3	335	1	GCP_AQUUB
30	515.5	24.3	346	2	07VKN4
31	512	24.1	364	2	06GIR3

32	510.5	24.0	346	2	07W668	07W668 bordetella
33	510	24.0	339	2	06LV10	06LV10 photobacter
34	509.5	24.0	346	2	07W134	07W134 bordetella
35	507.5	23.9	357	2	089WM1	089WM1 bradyrhizob
36	503.5	23.7	341	2	07NUE3	07NUE3 chromobacte
37	494.5	23.3	343	2	09CLJ1	09CLJ1 pasteurella
38	494	23.2	341	2	08ES16	08ES16 oceanobacil
39	492.5	23.2	337	1	GCP_SALTY	094731 salmonella
40	492.5	23.2	341	2	0915V7	0915V7 pseudomonas
41	491	23.1	364	2	06FYF1	06FYF1 bartonella
42	490.5	23.1	337	2	08FDG6	08FDG6 escherichia
43	489	23.0	353	2	07WQD9	07WQD9 candidatus
44	488.5	23.0	337	2	08XBK3	08XBK3 escherichia
45	486.5	22.9	337	2	082XN2	082XN2 nitrosomona

## ALIGNMENTS

RESULT 1	ID	096EV9	PRELIMINARY:	PRT:	414 AA.
AC	096EV9				
DT	01-DEC-2001 (TREMBLrel. 19, Created)				
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)				
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)				
DE	OSGEP1 protein.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Muscle;				
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;				
RA	Strausberg R.L., Peingold E.A., Grouse L.H., Derge J.G.,				
RA	Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schaller G.D.,				
RA	Altschul S.F., Zeeberg B., Buecaw K.H., Scheffer C.F., Bhat N.K.,				
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,				
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,				
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,				
RA	Brownstein M.J., Ueda N., Toshiyuki S., Carninci P., Prange C.,				
RA	Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,				
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunnarsson P.H.,				
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,				
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA	Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,				
RA	Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,				
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,				
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,				
RA	Krzyszinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,				
RA	Jones S.J., Marra M.A.,				
RT	"Generation and initial analysis of more than 15,000 full-length human				
RT	and mouse cDNA sequences."				
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Muscle;				
RA	Strausberg R.L.				
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; BC011904; AAH11904.1; -				
DR	MEROPS; M2.004; -				
DR	GO; GO:0008450; F-O-gialoglycoprotein endopeptidase activity; IEA.				
DR	GO; GO:0008270; F-zinc ion binding; IEA.				
DR	GO; GO:000508; P-proteolysis and peptidolysis; IEA.				
DR	InterPro; IPR000905; Peptidase M22.				
DR	InterPro; IPR009180; Pept M22_Osialgl.				
DR	Pfam; PF00814; Peptidase M22; 1.				
DR	PIRSF; PIRSF004537; Osialgl_pttda; 1.				
DR	PRINTS; PR00789; OSIALOPTASE.				
DR	ProDom; PD002367; Peptidase M22; 1.				
DR	TIGRFAMs; TIGR00329; GCP; 1.				
DR	SEQUENCE 414 AA; 45122 MW; A536B333F5C6B8DD CRC64;				

Query Match 100.0%; Score 2125; DB 2; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-155;  
 Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITKTAGVFPKSKRKYEFLLRSFNHPGTLFLHKLIVLGIEISCDPTAAAVDETGNV 60  
 DB 1 LGEAHSQTEVHLKTKGIVPPAAQOHLHRENIQRIVOELASAGVSPDLSAATTIKPGL 120  
 QY 61 LGEAHSQTEVHLKTKGIVPPAAQOHLHRENIQRIVOELASAGVSPDLSAATTIKPGL 120  
 DB 61 LGEAHSQTEVHLKTKGIVPPAAQOHLHRENIQRIVOELASAGVSPDLSAATTIKPGL 120  
 QY 121 ALSGVLSFSLQVLGKLPPIPIHMEAHALTRLTNKVEFPFLVLLISGHCILALV 180  
 DB 121 ALSGVLSFSLQVLGKLPPIPIHMEAHALTRLTNKVEFPFLVLLISGHCILALV 180  
 QY 181 QGVSDPFLLGKSLDIAFGMDLVKVARLSLIGHPECSWGGKALEHLAKQGRFHPDIK 240  
 DB 181 QGVSDPFLLGKSLDIAFGMDLVKVARLSLIGHPECSWGGKALEHLAKQGRFHPDIK 240  
 QY 241 PPLHAKKCDPSEFTGLQHTVDKIIMKKEKEBEGIEKQILSSAADIAATVQHTMACHLYKR 300  
 DB 241 PPLHAKKCDPSEFTGLQHTVDKIIMKKEKEBEGIEKQILSSAADIAATVQHTMACHLYKR 300  
 QY 301 THRALFCCKORDLPONNAVIVASGVASNFYIRRALILTNAQCTLLCPPLCTDNG 360  
 DB 301 THRALFCCKORDLPONNAVIVASGVASNFYIRRALILTNAQCTLLCPPLCTDNG 360  
 QY 361 IMIANGIERLRAIGILHDIEGIRYBPKPLGVDISKEVGEASIKVPOLKMEI 414  
 DB 361 IMIANGIERLRAIGILHDIEGIRYBPKPLGVDISKEVGEASIKVPOLKMEI 414

RESULT 2

Q9H4B0 PRELIMINARY; PRT; 439 AA.

AC 09H4B0;  
 DT 01-MAR-2001 (TRENBLREL. 16, Created)  
 DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)  
 DT 01-MAR-2004 (TRENBLREL. 26, Last annotation update)  
 DE Putative sialoglycoprotein type 2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Uterus;  
 RA Chen J.M., Fortunato M., Barrett A.J.;  
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AJ295148; CAC14666.1; --  
 DR MEROPS; M22.004; --  
 DR Genew; HGNC:23075; OSGEPL1.  
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008223; F:peptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000905; Peptidase M22.  
 DR InterPro; IPR009180; Pept M22 Osa1gl.  
 DR Pfam; PF00814; Peptidase M22; 1.  
 DR PIRSF; PIRSF004537; Osa1glc\_ptide; 1.  
 DR PRINTS; PR00767; Osa1LOPTNSE.  
 DR ProDom; PD002367; Peptidase M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 KW Peptase.  
 SQ SEQUENCE 439 AA; 48040 MW; 44849372C784E41P CRC64;

Query Match 98.4%; Score 2090.5; DB 2; Length 439;  
 Best Local Similarity 93.8%; Pred. No. 1.3e-152;  
 Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLITKTAGVFPKSKRKYEFLLRSFNHPGTLFLHKLIVLGIEISCDPTAAAVDETGNV 60

DB 1 MLITKTAGVFPKSKRKYEFLLRSFNHPGTLFLHKLIVLGIEISCDPTAAAVDETGNV 60  
 QY 61 LGEAHSQTEVHLKTKGIVPPAAQOHLHRENIQRIVOELASAGVSPDLSAATTIKPGL 120  
 DB 61 LGEAHSQTEVHLKTKGIVPPAAQOHLHRENIQRIVOELASAGVSPDLSAATTIKPGL 120  
 QY 121 ALSGVLSFSLQVLGKLPPIPIHMEAHALTRLTNKVEFPFLVLLISGHCILALV 180  
 DB 121 ALSGVLSFSLQVLGKLPPIPIHMEAHALTRLTNKVEFPFLVLLISGHCILALV 180  
 QY 181 QGVSDPFLLGKSLDIAFGMDLVKVARLSLIGHPECSWGGKALEHLAKQGRFHPDIK 240  
 DB 181 QGVSDPFLLGKSLDIAFGMDLVKVARLSLIGHPECSWGGKALEHLAKQGRFHPDIK 240  
 QY 241 PPLHAKKCDPSEFTGLQHTVDKIIMKKEKEBEGIEKQILSSAADIAATVQHTMACHLYKR 300  
 DB 241 PPLHAKKCDPSEFTGLQHTVDKIIMKKEKEBEGIEKQILSSAADIAATVQHTMACHLYKR 300  
 QY 301 THRALFCCKORDLPONNAVIVASGVASNFYIRRALILTNAQCTLLCPPLCTDNG 360  
 DB 301 THRALFCCKORDLPONNAVIVASGVASNFYIRRALILTNAQCTLLCPPLCTDNG 360  
 QY 361 IMIANGIERLRAIGILHDIEGIRYBPKPLGVDISKEVGEASIKVPOLKMEI 414  
 DB 361 IMIANGIERLRAIGILHDIEGIRYBPKPLGVDISKEVGEASIKVPOLKMEI 414

RESULT 3

Q96NH5 PRELIMINARY; PRT; 364 AA.

AC 096NH5;  
 DT 01-DEC-2001 (TRENBLREL. 19, Created)  
 DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)  
 DT 01-MAR-2004 (TRENBLREL. 26, Last annotation update)  
 DE Hypothetical protein FLJ30879.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
 RA Sekine M., Obaishi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
 RA Yamamoto J., Saito K., Kawai T., Isono Y., Nakamura Y., Nagahara K.,  
 RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
 RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Shimizu K.,  
 RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikawa E., Omura Y.,  
 RA Abe K., Kamihara K., Katsura N., Sato K., Tanikawa M., Yamazaki M.,  
 RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,  
 RA Tanai H., Kimeta M., Watanabe M., Hirose S., Chiba Y., Ishida S.,  
 RA Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hottu T., Kusano J.,  
 RA Kanohori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,  
 RA Togo Y., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
 RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Omori Y.,  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togeishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,

RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita K.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs";  
RL Nat. Genet. 36:40-45(2004).  
DR EMBL: AK055441; BAB70923.1; -.  
DR MEROPS: M22.004; -.  
DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO: GO:0008233; F:peptidase activity; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR000905; Peptidase\_M22.  
DR InterPro: IPR009180; Pept\_M22\_Osialgl.  
DR Pfam: PF00814; Peptidase\_M22; 1.  
DR PIRSF: PIRSF004537; Osialglc\_ptclds; 1.  
DR PRINTS: PR00789; OSIALOPTASE.  
DR ProDom: PD002367; Peptidase\_M22; 1.  
DR TrIRFams: TIGR00329; gcp; 1.  
KW Protease.  
SQ SEQUENCE 364 AA; 39528 MW; E0B605A07D0EC3D6 CRC64;  
  
Query Match 86.8%; Score 1845; DB 2; Length 364;  
Best Local Similarity 99.2%; Pred. No. 8,4e-134;  
Matches 361; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 MLITKTAGVFFPKSKRYEFLRSFNFHPTLFLHKIVLGIETSCDDTAAAVDEGNV 60  
Db 1 MLITKTAGVFFPKSKRYEFLRSFNFHPTLFLHKIVLGIETSCDDTAAAVDEGNV 60  
QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIQVEALSASGVSPSDLSAATTIKPGL 120  
Db 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIQVEALSASGVSPSDLSAATTIKPGL 120  
QY 121 ALSLGVLSFSLQVQLKKPFIPIHMEAHALITRLTNKVEPFLVLLISGHCLALV 180  
Db 121 ALSLGVLSFSLQVQLKKPFIPIHMEAHALITRLTNKVEPFLVLLISGHCLALV 180  
QY 181 QGVSDFLILGKSLDIAPGDMLDKVARRLSLIKHPECSMTSGSKAIEHLAKQGNRFHFDIK 240  
Db 181 QGVSDFLILGKSLDIAPGDMLDKVARRLSLIKHPECSMTSGSKAIEHLAKQGNRFHFDIK 240  
QY 241 PPLHNAKCDPSFTGLQHTDVKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYKR 300  
Db 241 PPLHNAKCDPSFTGLQHTDVKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYKR 300  
QY 301 THRALLFCQKRDLLPQNNAVLVASGVASNFYIRALTEILTNATQCTLLCPPRLCTDNG 360  
Db 301 THRALLFCQKRDLLPQNNAVLVASGVASNFYIRALTEILTNATQCTLLCPPRLCTDNG 360  
QY 361 IMTA 364  
Db 361 IMTA 364  
  
RESULT 4  
Q6PEB4 PRELIMINARY; PRT; 414 AA.  
AC Q6PEB4;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Hypothetical protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C26G II; TISSUE=Mammary tumor;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klauener R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang Y., Hsieh F.,  
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Uediri T.B., Toshiyuki S., Carninci P., Plange C.,  
RA Raha S.S., Locuelli N.A., Peters G.J., Abramson R.D., Mullanb S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek U.A., Guarnier P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz U., Myers R.W., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,  
RA Jones S.J., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C26G II; TISSUE=Mammary tumor;  
RA Strausberg R.;  
RL Submitted (SEP-2003) to the EMBL/Genbank/DBJ databases.  
DR EMBL: BC058172; AAHS8172.1; -.  
DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR000905; Peptidase\_M22.  
DR InterPro: IPR009180; Pept\_M22\_Osialgl.  
DR Pfam: PF00814; Peptidase\_M22; 1.  
DR PIRSF: PIRSF004537; Osialglc\_ptclds; 1.  
DR PRINTS: PR00789; OSIALOPTASE.  
DR ProDom: PD002367; Peptidase\_M22; 1.  
DR TrIRFams: TIGR00329; gcp; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 414 AA; 44962 MW; 0418AD93C6FEBEBCAE CRC64;  
  
Query Match 86.4%; Score 1835; DB 2; Length 414;  
Best Local Similarity 85.0%; Pred. No. 5,8e-133;  
Matches 352; Conservative 25; Mismatches 37; Indels 0; Gaps 0;  
  
QY 1 MLITKTAGVFFPKSKRYEFLRSFNFHPTLFLHKIVLGIETSCDDTAAAVDEGNV 60  
Db 1 MLITKTAGVFFPKSKRYEFLRSFNFHPTLFLHKIVLGIETSCDDTAAAVDEGNV 60  
QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIQVEALSASGVSPSDLSAATTIKPGL 120  
Db 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIQVEALSASGVSPSDLSAATTIKPGL 120  
QY 121 ALSLGVLSFSLQVQLKKPFIPIHMEAHALITRLTNKVEPFLVLLISGHCLALV 180  
Db 121 ALSLGVLSFSLQVQLKKPFIPIHMEAHALITRLTNKVEPFLVLLISGHCLALV 180  
QY 181 QGVSDFLILGKSLDIAPGDMLDKVARRLSLIKHPECSMTSGSKAIEHLAKQGNRFHFDIK 240  
Db 181 QGVSDFLILGKSLDIAPGDMLDKVARRLSLIKHPECSMTSGSKAIEHLAKQGNRFHFDIK 240  
QY 241 PPLHNAKCDPSFTGLQHTDVKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYKR 300  
Db 241 PPMQNAKCDPSFTGLQHTDVKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYKR 300  
QY 301 THRALLFCQKRDLLPQNNAVLVASGVASNFYIRALTEILTNATQCTLLCPPRLCTDNG 360  
Db 301 THRALLFCQKRDLLPQNNAVLVASGVASNFYIRALTEILTNATQCTLLCPPRLCTDNG 360  
QY 361 IMIANGIERLRAGGILHIDIGIRYEPKCPGVDSKEVGSASIVPOLKKEI 414  
Db 361 IMIANGIERLRAGGILHIDIGIRYEPKCPGVDSKEVGSASIVPOLKKEI 414  
  
RESULT 5  
Q8BLB6 PRELIMINARY; PRT; 414 AA.  
ID Q8BLB6



AC Q8BLB6; 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length  
 DE enriched library, clone: B20219017 product: similar to PUTATIVE  
 DE SIALOGLYCOPROTEINASE TYPE 2.  
 GN Name=Osgp11;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system: 384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hitozawa T.,  
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Konda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN EMBL; AK045669; BAC32450.1; --

DR MEROPS; M22.004; -;  
 DR MGD; MGI:1919335; Osgp11.  
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000905; Peptidase\_M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Osialglc\_ptcde; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR ProDom; PD002677; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 KW Protease.  
 SQ SEQUENCE 414 AA; 44927 MW; 9A2D27B576C7DB24 CRC64;  
 Query Match 86.0%; Score 1827; DB 2; Length 414;  
 Best Local Similarity 85.0%; Pred. No. 2,4e-132;  
 Matches 352; Conservative 23; Mismatches 39; Indels 0; Gaps 0;  
 QY 1 MLILFKTAGVFFKPSKRYEFLRSFNFPGTLFLHKIYVGIETSCDDTAAAVDEGTGV 60  
 DB 1 MMLMRTAGAIKPKPSKRYGFLRRFSVHPRLISCHLVGIETSCDDTGAAYVDGTGV 60  
 QY 61 LGEAHSQTEVALKTGGIVPPAQQIARENIQRIYQEAALSASGVSPDSLSAIATTKPGI 120  
 DB 61 LGEALHSQTEVALKTGGIVPPAQQIARENIQRIYQEAALSASGVSPDSLSAIATTKPGI 120  
 QY 121 ALSLVGVSFSLQVGLKRPPIPIHMEAAHLTRLTKNVEPFLVLLISGCHLALV 180  
 DB 121 ALSLVGVSFSLQVGLKRPPIPIHMEAAHLTRLTKNVEPFLVLLISGCHLALV 180  
 QY 181 QGVSPFLIGKSLDIARGMDLKVARRSLIHPBCSTSGGKAIHLAKOGRFPFDIK 240  
 DB 181 QGVSPFLIGKSLDIARGMDLKVARRSLIHPBCSTSGGKAIHLAKOGRFPFDIK 240  
 QY 241 PELHAKNCDFFSTGLOHTYTDKTIIMKEKEBGIKQILSSADIATVQHTWACHLVKR 300  
 DB 241 PMQNAKNCDFFSTGLOHTYTDKTIIMKEKEBGIKQILSSADIATVQHTWACHLVKR 300  
 QY 301 THRATLFCQKQMLSPANALVAVSGVSNFYIRALLETITNAQTCTLLCPPLCTDNG 360  
 DB 301 THRATLFCQKQMLSPANALVAVSGVSNFYIRALLETITNAQTCTLLCPPLCTDNG 360  
 QY 361 IMIANGIERLRAGGIIHDIGIRERPKCPGVDSKEVGEASIVPOLKMEI 414  
 DB 361 IMIANGIERLRAGGIIHDIGIRERPKCPGVDSKEVGEASIVPOLKMEI 414  
 RESULT 6  
 ID Q9DON0 PRELIMINARY; PRT; 414 AA.  
 AC Q9DON0;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Mus musculus 10 days embryo whole body cDNA, RIKEN full-length  
 DE enriched library, clone: 2610001M19 product: similar to PUTATIVE  
 DE SIALOGLYCOPROTEINASE TYPE 2.  
 GN Name=Osgp11;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Whole body;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA RIKEN FANTOM Consortium;  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Whole body;  
RA the FANTOM Consortium,  
RT the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Whole body;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/9r.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes."  
RL Genome Res. 10:1617-1630(2000).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Whole body;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/9r.152600;  
RA Shibata K., Itoh M., Alzawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Kono H., Akiyama Y., Nishi K., Kitsuami T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwara K., Inoue K., Ozawa Y., Izawa M., Ohara E., Watanabe M.,  
RA Yoneda Y., Ishikawa T., Tanaka K., Tanaka T., Matsura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer."  
RL Genome Res. 10:1757-1771(2000).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Whole body;  
RA Adachi T., Alzawa K., Akahira S., Akimura T., Arai A., Aono H.,  
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hirooka T., Hori F.,  
RA Inotani K., Ishii Y., Itoh M., Izawa M., Kankawa T., Kato H.,  
RA Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurahara C.,  
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ono M.,  
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
RA Muramatsu M., Hayashizaki Y.,  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
RN EMBL: AK011265; BAB27506.1; -  
DR MEROPS; M22.004; -  
DR MGD; MGI:191333; Osegep1.  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008233; F:Peptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR000905; Peptide\_M22.  
DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
DR Pfam; PF00814; Peptidase\_M22; 1.  
DR PIRSF; PIRSF004537; Osialglc\_ptids; 1.  
DR PRINTS; PR00789; OSIALOPTASE.  
DR ProDom; PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs; TIGR00329; gcp; 1.  
KM Protease.  
SO SEQUENCE 414 AA; 44999 MW; 999BC689944DDB24 CRC64;

Db 1 MMLRRYAGIIPPPKSKYGFRRPSVHPRTLSCHKLVIGITSCDDTAAAVDEGNV 60  
QY 61 LGEAHSQTEVHLKGTGIIVPPAAQOLHRENIQRIQVQALASASGSPDSALATTKPGL 120  
Db 61 LGEAHSQTEVHLKGTGIIVPPAAQOLHRENIQRIQVQALASASGSPDSALATTKPGL 120  
QY 121 ALSLVGLSPSLQVLQKPKPIPIHMEAAHTIRLTNRKVEPPVLLISGHCULALV 180  
Db 121 ALSLVGLSPSLQVLQKPKPIPIHMEAAHTIRLTNRKVEPPVLLISGHCULALV 180  
QY 181 QGVSDPDLGLKSLDIPAGMDLVNARLSIKRPESTSGGAIEHLAAGNRRFFPDK 240  
Db 181 QGVSDPDLGLKSLDIPAGMDLVNARLSIKRPESTSGGAIEHLAAGNRRFFPDK 240  
QY 241 PLTHAKNCDFSPFGQHTYTDKTIKKEKEEGEKGQIISADIAVQHTMACHLVK 300  
Db 241 PLTHAKNCDFSPFGQHTYTDKTIKKEKEEGEKGQIISADIAVQHTMACHLVK 300  
QY 301 THRALLFCRKORDLPPONAVLVASGVASNFYIRRALETLTNTACTLLCPPLCTDNG 360  
Db 301 THRALLFCRKORDLPPONAVLVASGVASNFYIRRALETLTNTACTLLCPPLCTDNG 360  
QY 361 IMIANGIRLRAGLGLHDIGIRYEPKCPICVDISKEVGEASIVPOLKMEI 414  
Db 361 IMIANGIRLRAGLGLHDIGIRYEPKCPICVDISKEVGEASIVPOLKMEI 414  
RESULT 7  
QY 06AVN7 PRELIMINARY; PRT; 467 AA.  
AC 06AVN7;  
DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Hypothetical protein (Fragment).  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxId=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins P.S., Wagner L., Shemen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F.,  
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Uedlin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Rana S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kerteman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,  
RA Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Kravinsky M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,  
RT Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences."  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX Director KGC Project;  
RL Submitted (Aug-2004) to the EMBL/GenBank/INP databases.  
DR EMBL; BC078974; AAH78974.1; -  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

```
DR InterPro; IPR000905; Peptidase_M22.
DR InterPro; IPR009180; Pept_M22_Osialgl.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00814; Peptidase_M22; 1.
DR PIRSF; PIRSF004537; Osialglc_ptlds; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR ProDom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGR00329; gcp; 1.
DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
KW Hypothetical protein.
FT NON_TER 467
SQ SEQUENCE 467 AA; 50799 MW; 474E18B1959B8AC0 CRC64;

Query Match      81.2%; Score 1725; DB 2; Length 467;
Best Local Similarity 86.1%; Pred. No.2e-124;
Matches 335; Conservative 19; Mismatches 35; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFFKSKRYEFLRSFNHPGLFLAKIVLGIEISCDDTAAAVDEIGNV 60
DB 1 MLMSTKTAGAIPRRPRSNVRGPIRRFNVQPALFHHKLVIGIETSCDDTAAAVDEIGNV 60
QY 61 LGEAIIHSQTEVHLKTGGIVPPAQQQIHRNTQRIYQELASGVSPPSLSAITTKRGL 120
DB 61 LGEAIIHSQTEVHLKTGGIVPPAQQQIHRNTQRIYEBALASGVSPPSLSAITTKRGL 120
QY 121 ALISGVGLSPSLQVGO;KKRPPIPIHMEAHALITRLNNKVEPPLVLLISGHCILALV 180
DB 121 ALISGVGLSPSLQVGO;KKRPPIPIHMEAHALITRLNNKVEPPLVLLISGHCILALV 180
QY 181 QGVSDPFLIGKSLDIPAGDMLDKVARRLSLIKHPECSMTSGGKALEHLAKQGNRPHFDIK 240
DB 181 QGVSDPFLIGKSLDIPAGDMLDKVARRLSLIKHPECSMTSGGKALEHLAKQGNRPHFDIK 240
QY 241 PPLHAAKNCDFSTGLQHVTDKTIKKKEKEGIEKGQILSSAADIATVQHTMACHLYKR 300
DB 241 PPMQNAKNCDFSTGLQHVTDKTIKKKEKEGIEKGQILSSAADIATVQHTMACHLYKR 300
QY 301 THRATLIFCQKQDILPQNNNAVIVASGVASNPFIYRALILTNATQCTLLCPPLCTONG 360
DB 301 THRATLIFCQKQDILPQNNNAVIVASGVASNPFIYRALILTNATQCTLLCPPLCTONG 360
QY 361 IMIANGIERLRAGLGLHDIEGIRYEPK 389
DB 361 IMIANGIERLRAGLGLHDIEGIRYEPK 389

RESULT 8
Q8JFW3 PRELIMINARY; PRT; 404 AA.
AC Q8JFW3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE SI:d221013.4 (Novel glycoprotease).
GN Name=d22B14.6;
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Babage A.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL591593; CAD4347.1; -.
DR MEROPS; M22.004; -.
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR00905; Peptidase_M22.
DR InterPro; IPR009180; Pept_M22_Osialgl.
DR Pfam; PF00814; Peptidase_M22; 1.
```

```
DR PIRSF; PIRSF004537; Osialglc_ptlds; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR ProDom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGR00329; gcp; 1.
KW Protease.
SQ SEQUENCE 404 AA; 43956 MW; 3A6B1DD050737B35 CRC64;

Query Match      60.4%; Score 1283; DB 2; Length 404;
Best Local Similarity 64.4%; Pred. No.1.9e-90;
Matches 242; Conservative 58; Mismatches 76; Indels 0; Gaps 0;

QY 37 KIVLGIEISCDDTAAAVDEIGNVIGEAIIHSQTEVHLKTGGIVPPAQQQIHRNTQRIYQ 96
DB 25 RLVLGIEISCDTEGAVALDETERIIGESIHSQETHLKTGGIIPVAQRHRENTSRVQQ 84
QY 97 EALASGVSPSLSAITTKRGLALISGVGLSPSLQVGO;KKRPPIPIHMEAHALITR 156
DB 85 EALASGVSPSLSAITTKRGLALISGVGLSPSLQVGO;KKRPPIPIHMEAHALITR 144
QY 157 LTNKVEPPLVLLISGHCILALVGVSDPFLIGKSLDIPAGDMLDKVARRLSLIKHPEC 216
DB 145 MHPIDPPLVLLISGHCILALVAGIDFLLGQTLDBAADTLDKIARRLSLRNHPEC 204
QY 217 STMGGKALEHLAKQGNRPHFDIKPPLHAAKNCDFSTGLQHVTDKTIKKKEKEGIEKG 276
DB 205 GTLSCGQAIERLAKEGDRLAFHIFSPMGQNYDCNFSFAGLRQITGAINKKEKEGIEKG 264
QY 277 QILSSAADIATVQHTMACHLYKTRATLIFCQKQDILPQNNNAVIVASGVASNPFIYRRA 336
DB 265 QPLSCVCKDIAAASQHTVASHLAKRTRATLIFCQKQDILPQNNNAVIVASGVASNPFIYRRI 324
QY 337 LEILTNATQCTLLCPPLCTONGIMIANNGIERLRAGLGLHDIEGIRYEPKPLGVDI 396
DB 325 LKTIITDATALHLCPSPSKCTONGIMIANNGIERLRAGLGLHDIEGIRYEPKPLGVDI 384
QY 397 SKEVGEASIKVPOLYK 412
DB 385 TSEVKEAAIKVPOLYK 400

RESULT 9
Q8JFW7 PRELIMINARY; PRT; 404 AA.
AC Q8JFW7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE SI:d272B14.6 (Novel glycoprotease).
GN Name=SI:d272B14.6;
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Clark G.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL672217; CAD43443.1; -.
DR MEROPS; M22.004; -.
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR00905; Peptidase_M22.
DR InterPro; IPR009180; Pept_M22_Osialgl.
DR Pfam; PF00814; Peptidase_M22; 1.
DR PIRSF; PIRSF004537; Osialglc_ptlds; 1.
DR PRINTS; PR00789; OSIALOPTASE.
DR ProDom; PD002367; Peptidase_M22; 1.
DR TIGRFAMs; TIGR00329; gcp; 1.
DR Protease.
SQ SEQUENCE 404 AA; 44027 MW; 6FB98653A651860F CRC64;
```

Query Match	60.2%	Score 1279	DB 2	Length 404
Best Local Similarity	64.1%	Pred. No. 3.9e-90		
Matches 241	Conservative 59	Mismatches 76	Indels 0	Gaps 0
QY	37	KIVLIERFSCDDTAAVVDETSNVLGEALHSOTVEHLKTGGIVPPAAQOLHRENIQRIYO	96	
DB	25	RUVIGIFRISCDTGAADVDETRILIGESLSHOKETHLTKGILIPVAQGLHRENIQRVQ	84	
QY	97	EALSAGVSPSOLSAIATTIKPGLALSLGVGSFSLIQVGLKKPEPIPIHMEAHALTR	156	
DB	85	EALNSAIEPSELTVAVATVTKGALSLGIGDYSILKFRQHQKPEPIPIHMEAHALTR	144	
QY	157	LTNKPEPPVVLVLIIGGCHLTLVGVSDPFLILGSLDIAFDMDLKVARRSLIKHPC	216	
DB	145	MLHPIDPFPLVVLVVGSHSLALAGLIDBFLLIGQLTDEAADPTLDKIKRRSLRNHPEC	204	
QY	217	STMSGCKAI EHLAKXGNRPHFDIKPPLHNAKNCDSFSTGLQHTYDKI IKKKEKESEIENG	276	
DB	205	GTLSGGQILIERLAKGGDOLAFHIFIPMGQNYCNCSFAGLRQITGAILKKKEKESEVEXG	264	
QY	277	QILSSADIAATVQHTMACHLVKTRHAILFCQRDLFPONNAVIVASGVASNPYIRRA	336	
DB	265	QFLSCVCKDIAAASQHTVASHLAKTRHAILFCCKSGILPEQNPPTLIVSGVASNETYIRI	324	
QY	337	LEILTNATQCTLLCEPPRLCTDNGIMLAWNGIERLPAGIGLIHIDIEGIEYEKCPGLVDI	396	
DB	325	LKIITDAGLHLLCPSPKCTCTDNGVIMWNGIERLKOEGKILSYSEEVSEYEPKAPGLDI	384	
QY	397	SKVEGASIKVPOLKM 412		
DB	385	TSEVEKAALIKVPKXL 400		
RESULT 10				
QYQ918				
ID	Q7Q918	PRELIMINARY	PRT	401 AA.
AC	Q7Q918			
DT	01-MAR-2004	(TREMBLrel. 26, Created)		
DT	01-MAR-2004	(TREMBLrel. 26, Last sequence update)		
DT	01-MAR-2004	(TREMBLrel. 26, Last annotation update)		
DE	AGCP14990	(Fragment)		
GN	Name=agCG46164; ORFNames=ENSANGC00000007922;			
OS	Anopheles gambiae ser. PEST.			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.			
OX	NCBI_TaxId=180454;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PEST;			
RA	Anopheles Genome Sequencing Consortium;			
RL	Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.			
CC	-!- CAUTION: The sequence shown here is derived from an			
CC	EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is			
CC	preliminary data.			
DR	EMBL; AAA01008900; EAA09387.1; -			
DR	GO; GO:0008450; F-O-sialoglycoprotein endopeptidase activity; IEA.			
DR	GO; GO:0008270; F:zinc ion binding; IEA.			
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	InterPro; IPR009180; Pept_M22_Ost191.			
DR	Pfam; PF00814; Peptidase_M22; 1.			
DR	PIRSF; PIRSF004537; Ostialytic_ptds; 1.			
DR	PRINTS; PR00789; OSTALOPTASB			
DR	ProDom; PD002367; Peptidase_M22; 1.			
DR	TIGRFAMs; TIGR00329; gcp; 1.			
FT	NON TER 1			
SO	SEQUENCE 401 AA; 43849 MW; CC9426723D3PD4F1 CRC64;			
Query Match	36.4%	Score 772.5	DB 2	Length 401
Best Local Similarity	42.7%	Pred. No. 4.2e-51		
Matches 165	Conservative 68	Mismatches 142	Indels 11	Gaps 4

QY	38	YLGLTSCDDPRAA>VNDSTGNVIGAHISQTEVHLKTGGLVPPAAOQLRENTQRLVOE	97
Db	9	VVLGJLSTSCDDPGLALVTONGVNLGSGYIHSQOSSHLPFGGIIIPVADIIHPANIESVON	68
QY	98	ALSAGVSPSDLSAIAITTKPGIALSLGVGLSEFSLQVGLKTPFPIHHMEAHATIRL	157
Db	69	AKFLANMTENDIDAVAVTNRPGILPUSLIQGMRAKHIIASVYNPLFIHHQNALMAARM	122
QY	158	TNKVEPPLVLLISGGHCLLALVQGVSDPLLKSLIDIAFGDMLDKVARRISLHKPCS	217
Db	129	TSTIPEPFLCILLVSGSHSLLVFVESYARFRLLETLDLDDAGEALDKIARLKLJNVAKYA	188
QY	218	TMSGKALIEHLAKQG-----NFHPDIKXPLHAKNCDPEFTGLQHTYDKIIMKEKEBG	272
Db	189	QMSGQDQALPAAAOQAGKDTSAVEFPL--PLSKYRCQOFSEFAGLKNATRHILIERESTLH	246
QY	273	IEKGQILSSAADIATVQHTMACHLVKRTTHRALIFCKORPLT--PQNNAVLVASGVAASN	333
Db	247	LAPDALLPPEYAFCAFCFLKGVTRHMLHRTQRALEYERRKLPSDAEHRSLVSGVACN	306
QY	331	FYIRALBELITNATQCTLLCPPEPLCTDNGIMIANWGIERLRA--GLGILHDIEGIRYEP	388
Db	307	DVIFNALSSMAQOFGSVYTRPRPKLCTDNGTMIANMGEMKTLAKDTAEMTTKYEQVINDISG	366
QY	389	KCPLEVDLSKEVGEASIKVPOLKMEI 414	
Db	367	KCPIDBSLIDVKEANICKMAKAVDI 392	
RESULT 11			
QYVND6			
ID	Q9VND6	PRELIMINARY; PRT; 409 AA.	
AC	Q9VND6;		
DT	01-MAY-2000 (TREMBlrel. 13, Created)		
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)		
DT	01-MAR-2004 (TREMBlrel. 26, Last annotation update)		
DE	CGI4231-PA.		
GN	ORFNames=CGI4231;		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxId=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;		
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amaratunga C.P., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	Georgie R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.H., Blasei R.G., Champe M., Pfeiffer B.D.,		
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,		
RA	Abt J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,		
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,		
RA	Burtis J.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Dubin K.J., Evangelista C.C., Petraz C., Ferreira S., Fleischmann W.,		
RA	Gooder C., Garfield A.E., Gang N.S., Galbraith W.M., Glasser K.,		
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harey D., Heiman T.J., Hernandez J.R., Houck J.,		
RA	Hoskins D., Houston K.A., Howland T.J., Wei M.H., Idagawa C.,		
RA	Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA	Kimmel P.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		
RA	Laesle P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA	Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,		
RA	Metkulov G., Milbina N.V., Mobarry C., Morris J., Moshrefi A.,		
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,		
RA	Nelson D.R., Nelson K.A., Nixon K., Nussler D.R., Paciel J.M.,		
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,		



RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.T., Sakurai T., Satou M.,  
RA Seki M., Shim P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.,  
RL Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AC002387; AAB82636.2; -  
DR EMBL: AY024338; AAK00530.1; -  
DR EMBL: AY063864; AAL36220.1; -  
DR EMBL: AY117283; AAM51358.1; -  
DR PIR: E84888; E84888.  
DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO: GO:0008233; F:peptidase activity; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR009095; Peptidase\_M22.  
DR InterPro: IPR009180; Pept\_M22\_Osialgl.  
DR Pfam: PF00814; Peptidase\_M22; 1.  
DR PIRSF: PIRSF004537; Osialglc\_pptds; 1.  
DR PRINTS: PR00789; OSIALOPTASE.  
DR ProDom: PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs: TIGR00329; gcp; 1.  
KM Complete proteome.  
SQ SEQUENCE 480 AA; 5295 MW; 20DD6A86AC1FPAD CRC64;

Query Match 31.5%; Score 668.5; DB 2; Length 480;  
Best Local Similarity 39.4%; Pred. No. 5.4e-43;  
Matches 154; Conservative 60; Mismatches 130; Indels 47; Gaps 7;

QY 38 IVLGIEISCDPTAAAVNDETNVIGEALHSQTEVHLKGGIVPPAAQQLHRENIQRYO 97  
DB 85 VVLGIETSCDDTAAAVNDETNVIGEALHSQTEVHLKGGIVPPAAQQLHRENIQRYO 144  
QY 98 ALSASGSPSLSAIATIKPGALSLGVGSFSLQVLGQLKPKPPIPIHMEHAHALTR 157  
DB 145 ALDQANLTERKLSAVATVITGPGSLCLKVGKARKARVANGNSPLVGVHMEHAHALVRL 204  
QY 158 T-NKVEPFPVLVLLISGHCILLAVOGVSDPFLGKSLDAPGMDLKVARRSLIKHPBC 216  
DB 205 VEGELSPFPMALLISGHNLLVLAHKLGQYQLGTTVDALGEAVDKAKMLGLDMH--- 261  
QY 217 STMSGGKAIEHLAKQGNRFHFDIKRPLHAKNCDSFPTGLQHTVDTKIMKKEKEGIEKG 276  
DB 262 --RSGGPAVEBLLEGDAKSKYKFNVPKMYKDCNFSYAGLTKQVRLALEAKE----IDAK 315  
QY 277 QILSSA-----ADIAATVQHTWACHLVKRTIRALIFCKQRDLTPQNNALVASSGVA 328  
DB 316 CPVSSANTEDRRNRADIAASFQVAVLHLEKCEBAIDWALE--LEPSIKMTVSSGVA 372  
QY 329 SNFYIRBALILTNATQCTLLCPPERLCTDNGIMIANNGIERLRAGLGLHDIGIRYEPK 387  
DB 373 SNKYVRLRLNNIVENKMLKLVCPPEPSLCTDNGVMVAMWGLGHEFRVG-----R'DP 422  
QY 388 -----PKCPGLGVDSKEVGEA 403  
DB 423 PPAATPEPDVYVDRPRWPLGEEYAKGRSEA 453

RESULT 13  
Q73H71 PRELIMINARY; PRT; 335 AA.  
AC Q73H71;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Peptidase, M22 family protein.  
GN OrderedLocustNames=MD0699;  
OS Rickettsia pipientis wmel.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Rickettsiaceae; Wolbachiae; Wolbachia.  
QY NCBI\_TaxID=55077;  
RN  
RP SEQUENCE FROM N.A.  
RX PubMed=15024419;

RA Wu M., Sun L.V., Vamathavan J.J., Riegler M., Deboy R.T.,  
RA Brownlie J.C., McDraw E.A., Martin W., Esser C., Almadhinejad N.,  
RA Wiegand C., Madupu R., Beanan M.J., Brinkac L.M., Daugherty S.C.,  
RA Durkin A.S., Kolonay J.F., Nelson W.C., Mohamoud Y., Lee P.,  
RA Berry K.J., Young M.B., Ultebrack T.R., Weidman J.F., Nierman W.C.,  
RA Paulsen I.T., Nelson K.E., Tettelin H., O'Neill S.L., Eisen J.A.,  
RT "Phylogenomics of the reproductive parasite Wolbachia pipientis wmel:  
RT a streamlined genome overrurn by mobile genetic elements."  
RL Plos Biol. 2:327-341(2004).  
DR EMBL: AE017258; AAS14395.1; -  
DR TIGR: WP0699; -  
DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR009095; Peptidase\_M22.  
DR InterPro: IPR009180; Pept\_M22\_Osialgl.  
DR Pfam: PF00814; Peptidase\_M22; 1.  
DR PIRSF: PIRSF004537; Osialglc\_pptds; 1.  
DR PRINTS: PR00789; OSIALOPTASE.  
DR ProDom: PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs: TIGR00329; gcp; 1.  
KM Complete proteome.  
SQ SEQUENCE 335 AA; 36634 MW; F4914CDA5B8C9390 CRC64;

Query Match 27.2%; Score 577.5; DB 2; Length 335;  
Best Local Similarity 36.3%; Pred. No. 3.6e-36;  
Matches 128; Conservative 68; Mismatches 130; Indels 27; Gaps 6;

QY 37 KIVLGIEISCDPTAAAVNDETNVIGEALHSQTEVHLKGGIVPPAAQQLHRENIQRYO 96  
DB 2 KTLIAVETSCDDTAAAVNDETNVIGEALHSQTEVHLKGGIVPPAAQQLHRENIQRYO 60  
QY 97 BALSASGSPSLSAIATIKPGALSLGVGSFSLQVLGQLKPKPPIPIHMEHAHALTR 156  
DB 61 SAVERSNLNFCDMLNLAATSGELIGLIVGTMAAIAHVAQKPIAVAHLEHAHALVLR 120  
QY 157 LTNKVEPFPVLVLLISGHCILLAVOGVSDPFLGKSLDAPGMDLKVARRSLIKHPBC 216  
DB 121 LHHEVKEPFPVLVLLISGHCILLAVOGVSDPFLGKSLDAPGMDLKVARRSLIKHPBC 174  
QY 217 STMSGGKAIEHLAKQGNRFHFDIKRPLHAKNCDSFPTGLQHTVDTKIMKKEKEGIEKG 276  
DB 175 --SYPGPPLLEKAKKNGNTRFPLKRAMIKRSGCNFSFGIKTAVKMLVDELKMS----- 228  
QY 277 QILSSAADIAATVQHTWACHLVKRTIRALIFCKQRDLTPQNNALVASSGVA SNFYIRBA 336  
DB 229 ---QVDCVCA SFQECISIDILDRVSNALIMESLNIKIND---FYITGVANANFLREK 282  
QY 337 LEILTNATQCTLLCPPERLCTDNGIMIANNGIERLRAGLGLHDIGIRYEPK 389  
DB 283 LKQHN---LNIFFPPNDLCTDNALMVGMTGIERLQKNT-----IDPLNFAFR 327

RESULT 14  
Q92LH8 PRELIMINARY; PRT; 360 AA.  
AC Q92LH8;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE PROBABLE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57).  
GN ORFNames=SKC03230;  
OS Rickettsia melliloti (Sinorhizobium melliloti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RA MEDLINE=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;  
RA Capela D., Barloy-Hubler F., Couzy J., Bothe G., Ampe F., Batut J.,  
RA Boistard P., Becker A., Boutry M., Cadieu B., Dreano S., Gloux S.,  
RA Godrie T., Goffeau A., Kahn D., Kiss B., Lelaure V., Masuy D.,



RA Pohl T., Portetelle D., Puehler A., Fumelle B., Ramsperger U.,  
 RA Remard C., Thebault P., Vandenbol M., Weidner S., Gallbert F.,  
 RT "Analysis of the chromosome sequence of the legume symbiont  
 RT Sinorhizobium meliloti strain 1021."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
 DR EMBL, AL591792; CAC47657.1; -.  
 DR GO, GO:0016787; F:Hydrolase activity; IEA.  
 DR GO, GO:0008450; F:O-6-ialoglycoprotein endopeptidase activity; IEA.  
 DR GO, GO:0008270; F:zinc ion binding; IEA.  
 DR GO, GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR000905; Peptidase M22.  
 DR InterPro: IPR00180; Pept. M22\_Ostalg1.  
 DR Pfam, PF00814; Peptidase\_M22; 1.  
 DR PIRSF, PIRSF004537; Ostalgic\_ptcds; 1.  
 DR PRINTS, PR00789; OSIALOPTASB.  
 DR Prodom: PD002367; Peptidase\_M22; 1.  
 DR TIGRPFAMs, TIGR00329; gcp; 1.  
 KW Complete proteome; Hydrolyase.  
 SQ SEQUENCE 360 AA; 37906 MW; A07F946AB562EA86 CRC64;  
 Query Match 26.8%; Score 568.5; DB 2; Length 360;  
 Best Local Similarity 37.6%; Pred. No. 1.9e-35;  
 Matches 143; Conservative 57; Mismatches 133; Indels 47; Gaps 9;  
 QY 39 VLGIETSCDDTAAAVV--DE--TGNVLGEAIIHSQTEVHLKTGIVPPAAQQLHRENIORI 94  
 DB 3 ILGIETSCDETAASVVALDEGRGRILGDVLSQLEHSAYGVVPEIARAHEALDGL 62  
 QY 95 VOELASAGVSPSDLSAIAITTIKPLALSLVGLSFLQVQLKKPFIPIHMEAHALT 154  
 DB 63 IEEALLRAGVTLRIDAAVATSGPGLIGLIVGLMTGKAIARATKPLVAVNHLGHALT 122  
 QY 155 IRLTNKVEPPLVLLISGGHCLLALVOGVDFLLGKSIDIPAGMLDKVARRLSIKHP 214  
 DB 123 ARLTGSLERPYLLLVSGHQTOLIVKGVGERKGTITDLAGFAFDKAKMLGL-PPY 181  
 QY 215 ECSTMSGKALIEHLAKQGNRFHFDIKPPLHNAKNCDFSEFTGLQ-----HVTDKI 263  
 DB 182 -----GGPAAVERAQAQNAERFDFPRPLVGDARLDFSGKLTAVRQAQSLGPTDQD 235  
 QY 264 IMKKEKEBGEIKGQILLSSADIAATVQHTMACHLYKTRHAILFCRKORDLLPQNNNAVLYA 323  
 DB 236 I-----ADVCAISFORAISRTLDRVGRGLKRRF-ADFAVDOPALV 276  
 QY 324 SGVAASNYIRALEILTNATQCTLLCPPRPLCTDNGIMIANNGIERLAGIILHIDEG 383  
 DB 277 AGVAANQTLRRITLQSLCDEHGFRTAPLPQLCTDAAAMIAWAGABRLAAGL----PADG 332  
 QY 384 IRVEPKC--PLGVDISKEVG 401  
 DB 333 LDAAPRSRWPLDSEKALIG 352  
 RESULT 15  
 Q98E16 PRELIMINARY; PRT; 362 AA.  
 AC Q98E16.  
 DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE O-6-ialoglycoprotein endopeptidase (Gcp).  
 GN OrderedLocustNames=mlr4224;  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OC NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Matanabe A., Idekawa K., Ishikawa K., Kawashima M., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,

RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.,  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti."  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL, AP003003; BAB50932.1; -.  
 DR GO, GO:0008450; F:O-6-ialoglycoprotein endopeptidase activity; IEA.  
 DR GO, GO:0008270; F:zinc ion binding; IEA.  
 DR GO, GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR000905; Peptidase M22.  
 DR InterPro: IPR00180; Pept. M22\_Ostalg1.  
 DR Pfam, PF00814; Peptidase\_M22; 1.  
 DR PIRSF, PIRSF004537; Ostalgic\_ptcds; 1.  
 DR PRINTS, PR00789; OSIALOPTASB.  
 DR Prodom: PD002367; Peptidase\_M22; 1.  
 DR TIGRPFAMs, TIGR00329; gcp; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 362 AA; 37345 MW; 8E4B11C45895D1D7 CRC64;  
 Query Match 26.5%; Score 564; DB 2; Length 362;  
 Best Local Similarity 38.4%; Pred. No. 4.3e-35;  
 Matches 145; Conservative 51; Mismatches 148; Indels 34; Gaps 10;  
 QY 39 VLGIETSCDDTAAAVV----DETGNVLGEAIIHSQTEVHLKTGIVPPAAQQLHRENIORI 94  
 DB 4 VLGIETSCDETAASVVALDEGRGRILGDVLSQLEHSAYGVVPEIARAHEALDGL 63  
 QY 95 VOELASAGVSPSDLSAIAITTIKPLALSLVGLSFLQVQLKKPFIPIHMEAHALT 154  
 DB 64 VEAALADSGTTLADIDALIAATAGPGLVGLVGLMTAKAIIAAGAPLIAVNHLEHALT 123  
 QY 155 IRLTNKVEPPLVLLISGGHCLLALVOGVDFLLGKSIDIPAGMLDKVARRLSIKHP 214  
 DB 124 ARLTGSLERPYLLLVSGHQTOLIVKGVGERKGTITDLAGFAFDKAKMLGL-PPY 182  
 QY 215 ECSTMSGKALIEHLAKQGNRFHFDIKPPLHNAKNCDFSEFTGLQ-----HVTDKI 274  
 DB 183 -----GGPAAVERAQAQNAERFDFPRPLVGDARLDFSGKLTAVRQAQSLGPTDQD 230  
 QY 275 K-QQLISSADIAATVQHTMACHLYKTRHAILFCRKORDLLP-QNNNAVLYASGVASNFY 332  
 DB 231 PLGDD-QDVADICASQAADVADALADRVSRLARFSQ--TFPGTKPALVAVAGVANHT 286  
 QY 333 IRRALEILTNATQCTLLCPPRPLCTDNGIMIANNGIERLAGIILHIDEGIRVEPKCPL 392  
 DB 287 IKATLERLCAEAGFTIVAPPLKCTDNNAMIANAGIERLAGIILHIDEGIRVEPKCPL 343  
 QY 393 GVDISKEVGEASIKVPOL 410  
 DB 344 PLD-----SISAPMV 353  
 Search completed: November 10, 2005, 16:52:01  
 Job time: 110.35 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:51:52 ; Search time 33.3 Seconds  
(without alignments)  
928.069 Million cell updates/sec

Title: US-10-649-273-2  
Perfect score: 2125  
Sequence: 1 MLILTKTAGVFFPKSRKRVY.....DISKEVGASIKVPOLKMEI 414

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2125	100.0	414	US-10-067-443-2	Sequence 2, Appli
2	2090.5	98.4	439	US-10-067-443-19	Sequence 19, Appl
3	1385	65.2	267	US-10-067-443-22	Sequence 22, Appl
4	634	29.8	463	US-10-067-443-3	Sequence 3, Appli
5	549	25.8	179	US-10-067-443-25	Sequence 25, Appli
6	524	24.7	421	US-10-067-443-4	Sequence 4, Appli
7	524	24.7	421	US-10-067-443-28	Sequence 28, Appli
8	501	23.6	350	US-09-540-236-2726	Sequence 2726, Ap
9	499.5	23.0	401	US-09-252-991A-17372	Sequence 17372, A
10	488	23.0	342	US-08-087-797-3	Sequence 3, Appli
11	482.5	22.7	357	US-09-543-681A-6513	Sequence 6513, Ap
12	475.5	22.4	325	US-08-087-797-2	Sequence 2, Appli
13	474.5	22.3	343	US-09-489-039A-9221	Sequence 9221, Ap
14	472	22.2	363	US-09-107-532A-6609	Sequence 6609, Ap
15	461	21.7	336	US-08-987-121A-4	Sequence 4, Appli
16	457	21.5	327	US-10-067-443-5	Sequence 5, Appli
17	457	21.5	335	US-08-961-083-52	Sequence 52, Appli
18	457	21.5	335	US-09-536-784-52	Sequence 52, Appli
19	455	21.4	336	US-09-066-512-2	Sequence 2, Appli
20	454	21.4	366	US-09-134-000C-4956	Sequence 4956, Ap
21	453	21.3	336	US-09-583-110-4857	Sequence 4857, Ap
22	448	21.1	336	US-09-107-433-4221	Sequence 4221, Ap
23	443	20.8	368	US-09-134-001C-9909	Sequence 9909, Ap
24	433	20.4	341	US-03-145-624-2	Sequence 2, Appli
25	406.5	19.1	344	US-09-198-452A-213	Sequence 213, App
26	406.5	19.1	360	US-09-438-185A-196	Sequence 196, App
27	401.5	18.9	344	US-09-602-777A-148	Sequence 148, App

28	362	17.0	292	4	US-09-724-623-81	Sequence 81, Appl
29	359	16.9	214	4	US-09-328-352-4609	Sequence 4609, Ap
30	338.5	15.9	340	4	US-10-067-443-6	Sequence 6, Appli
31	319	15.0	273	4	US-09-710-279-728	Sequence 728, App
32	204	9.6	187	4	US-09-248-796A-19489	Sequence 19489, A
33	181.5	8.5	143	4	US-09-328-352-4387	Sequence 4387, Ap
34	109	5.1	515	4	US-09-252-991A-33046	Sequence 33046, A
35	100	4.7	225	4	US-09-602-777A-144	Sequence 144, App
36	97.5	4.6	580	4	US-09-489-039A-12013	Sequence 12013, A
37	97	4.6	601	3	US-09-155-036-5	Sequence 5, Appli
38	97	4.6	601	4	US-09-866-307-5	Sequence 5, Appli
39	97	4.6	614	3	US-09-866-307-6	Sequence 6, Appli
40	97	4.6	614	4	US-09-866-307-6	Sequence 6, Appli
41	96	4.4	612	4	US-09-902-540-13545	Sequence 13545, A
42	94.5	4.4	480	4	US-09-583-110-5050	Sequence 5050, Ap
43	94.5	4.4	481	4	US-09-107-433-3197	Sequence 3197, Ap
44	93.5	4.4	294	4	US-09-107-532A-6933	Sequence 6933, Ap
45	93.5	4.4	579	4	US-09-543-681A-5019	Sequence 5019, Ap

ALIGNMENTS

RESULT 1									
US-10-067-443-2									
; Sequence 2, Application US/10067443									
; Patent No. 6642041									
; GENERAL INFORMATION:									
; APPLICANT: Bristol-Myers Squibb Company									
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I									
; FILE REFERENCE: D0073 NP									
; CURRENT APPLICATION NUMBER: US/10/067,443									
; PRIOR FILING DATE: 2002-02-05									
; PRIOR APPLICATION NUMBER: US 60/266, 518									
; PRIOR FILING DATE: 2001-02-05									
; PRIOR APPLICATION NUMBER: US 60/282, 814									
; NUMBER OF SEQ ID NOS: 71									
; SOFTWARE: PatentIn version 3.0									
; SEQ ID NO 2									
; LENGTH: 414									
; TYPE: PRT									
; ORGANISM: Homo sapiens									
US-10-067-443-2									
Query Match									
Best Local Similarity 100.0%; Score 2125; DB 4; Length 414;									
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MLILTKTAGVFFPKSRKRVYEFLLRSFNHFGTLFLHKLVLGTFETSCDDPRAAVVDERTGV	60						
DB	1	MLILTKTAGVFFPKSRKRVYEFLLRSFNHFGTLFLHKLVLGTFETSCDDPRAAVVDERTGV	60						
QY	61	IGEAHSQTEVHLKTCGIVPPAAOOLHRENIORIVOEALASGVSPDSATATTKPGI	120						
DB	61	IGEAHSQTEVHLKTCGIVPPAAOOLHRENIORIVOEALASGVSPDSATATTKPGI	120						
QY	121	ALSLGVGSLFSQVLGVOLKKPPIPIHMEAHALLTRLTNKKVEPPLVLLISGHCILALV	180						
DB	121	ALSLGVGSLFSQVLGVOLKKPPIPIHMEAHALLTRLTNKKVEPPLVLLISGHCILALV	180						
QY	181	QCVSPFLGLKSLDIAFGMDLKVARRSLIKHPBCSTMSGKAIHHLAKGNRFHFDIK	240						
DB	181	QCVSPFLGLKSLDIAFGMDLKVARRSLIKHPBCSTMSGKAIHHLAKGNRFHFDIK	240						
QY	241	PLHLAKKNDSEFTGLQHTVTDKTIKKKEKESGILGQILSSAADIATVQHTMACHLVNR	300						
DB	241	PLHLAKKNDSEFTGLQHTVTDKTIKKKEKESGILGQILSSAADIATVQHTMACHLVNR	300						
QY	301	THRAILFCQKORLLPQNNNAVLAASGVASNFYIRRALILITVNAOCTLLCPPRLCTDNG	360						
DB	301	THRAILFCQKORLLPQNNNAVLAASGVASNFYIRRALILITVNAOCTLLCPPRLCTDNG	360						

QY 361 IMIANGIERLRAGIILHDIIEGIRYEPKCPGLGVDISKVEGASIKVPOLKMEI 414  
Db 361 IMIANGIERLRAGIILHDIIEGIRYEPKCPGLGVDISKVEGASIKVPOLKMEI 414

## RESULT 2

US-10-067-443-19  
Sequence 19, Application US/10067443

Patent No. 6642041

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED

FILE REFERENCE: D0073 NP

CURRENT APPLICATION NUMBER: US/10/067,443

CURRENT FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: Patentin version 3.0

SEQ ID NO 19

LENGTH: 439

TYPE: PRT

ORGANISM: homo sapiens

US-10-067-443-19

Query Match 98.4%; Score 2090.5; DB 4; length 439;  
Best Local Similarity 93.8%; Pred. No. 1.4e-230;

Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLITKTGVFPKSKRYEFLRSFNRHPTLFLHKIYLGJETSCTDTAAVVDETGV 60  
Db 1 MLITKTGVFPKSKRYEFLRSFNRHPTLFLHKIYLGJETSCTDTAAVVDETGV 60  
QY 61 LGEAHSQTEVHLKGTGIVPPAAQOLHRENIQIVQEAISASGVSPSDLSAATTIKPG 120  
Db 61 LGEAHSQTEVHLKGTGIVPPAAQOLHRENIQIVQEAISASGVSPSDLSAATTIKPG 120  
QY 121 ALSIGVGLSFLQVGVOLKPKPIPIHMEAHALTRITNKVEFPFLVLLISGHCILALV 180  
Db 121 ALSIGVGLSFLQVGVOLKPKPIPIHMEAHALTRITNKVEFPFLVLLISGHCILALV 180  
QY 181 QGVSDPFLLGKSLDIAPGMDLVKVARSLIHPPECSTWSGKAIEHLAKQGRFFHDIK 240  
Db 181 QGVSDPFLLGKSLDIAPGMDLVKVARSLIHPPECSTWSGKAIEHLAKQGRFFHDIK 240  
QY 241 PRLHAKKCDPFTGLQHTVDKTIIMKKEBEGI-----BK 275  
Db 241 PRLHAKKCDPFTGLQHTVDKTIIMKKEBEGI-----BK 275  
QY 276 GQISSAADIAATVOHTMACHLVKRTTHRAILFCQKQDLPONNAVVASGVASNFYIRR 335  
Db 276 GQISSAADIAATVOHTMACHLVKRTTHRAILFCQKQDLPONNAVVASGVASNFYIRR 335  
QY 301 GQISSAADIAATVOHTMACHLVKRTTHRAILFCQKQDLPONNAVVASGVASNFYIRR 360  
Db 301 GQISSAADIAATVOHTMACHLVKRTTHRAILFCQKQDLPONNAVVASGVASNFYIRR 360  
QY 336 ALEIITNATQCTLLCPPLCTDNGIMIAMNGIERLRAGIILHDIIEGIRYEPKCPGLVD 395  
Db 336 ALEIITNATQCTLLCPPLCTDNGIMIAMNGIERLRAGIILHDIIEGIRYEPKCPGLVD 395  
QY 361 ALEIITNATQCTLLCPPLCTDNGIMIAMNGIERLRAGIILHDIIEGIRYEPKCPGLVD 420  
Db 361 ALEIITNATQCTLLCPPLCTDNGIMIAMNGIERLRAGIILHDIIEGIRYEPKCPGLVD 420  
QY 396 ISKEVGEASIKVPOLKMEI 414  
Db 396 ISKEVGEASIKVPOLKMEI 414  
QY 421 ISKEVGEASIKVPOLKMEI 439  
Db 421 ISKEVGEASIKVPOLKMEI 439

## RESULT 3

US-10-067-443-22  
Sequence 22, Application US/10067443

Patent No. 6642041

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED

TITLE OF INVENTION: SPINAL CORD, MP-1  
FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
CURRENT FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 22  
LENGTH: 267  
TYPE: PRT  
ORGANISM: homo sapiens  
US-10-067-443-22

Query Match 65.2%; Score 1385; DB 4; length 267;  
Best Local Similarity 100.0%; Pred. No. 5.1e-150;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 MEHAALTITRTNKVEFPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLVKVAR 207  
Db 1 MEHAALTITRTNKVEFPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLVKVAR 207  
QY 208 LSLIKHPECSTWSGKAIEHLAKQGRFFHDIKPPLHAKKCDPFTGLQHTVDKTIIMK 267  
Db 208 LSLIKHPECSTWSGKAIEHLAKQGRFFHDIKPPLHAKKCDPFTGLQHTVDKTIIMK 267  
QY 268 EKEBEGIEKQIISADIAATVOHTMACHLVKRTTHRAILFCQKQDLPONNAVVASGV 327  
Db 268 EKEBEGIEKQIISADIAATVOHTMACHLVKRTTHRAILFCQKQDLPONNAVVASGV 327  
QY 328 ASNFYIRRALBITNATQCTLLCPPLCTDNGIMIAMNGIERLRAGIILHDIIEGIRY 387  
Db 328 ASNFYIRRALBITNATQCTLLCPPLCTDNGIMIAMNGIERLRAGIILHDIIEGIRY 387  
QY 388 PKCPLGVDISKVEGASIKVPOLKMEI 414  
Db 388 PKCPLGVDISKVEGASIKVPOLKMEI 414  
QY 414 PKCPLGVDISKVEGASIKVPOLKMEI 439  
Db 414 PKCPLGVDISKVEGASIKVPOLKMEI 439

## RESULT 4

US-10-067-443-3  
Sequence 3, Application US/10067443

Patent No. 6642041

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED

FILE REFERENCE: D0073 NP

CURRENT APPLICATION NUMBER: US/10/067,443

CURRENT FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: Patentin version 3.0

SEQ ID NO 3

LENGTH: 463

TYPE: PRT

ORGANISM: Arabidopsis thaliana

US-10-067-443-3

Query Match 29.8%; Score 634; DB 4; length 463;  
Best Local Similarity 37.9%; Pred. No. 1.8e-63;  
Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;

QY 38 IYVGITSCDDTAAVVDETGVNVLGEAHSQTEVHLKGTGIVPPAAQOLHRENIQIVQ 97  
Db 38 IYVGITSCDDTAAVVDETGVNVLGEAHSQTEVHLKGTGIVPPAAQOLHRENIQIVQ 97  
QY 85 VVGITETSCTDTAAVVDETGVNVLGEAHSQTEVHLKGTGIVPPAAQOLHRENIQIVQ 141  
Db 85 VVGITETSCTDTAAVVDETGVNVLGEAHSQTEVHLKGTGIVPPAAQOLHRENIQIVQ 141  
QY 98 ALSASGVSPSDLSAIAITTIKRGALSLGVGLSFLQVGVOLKPKPIPIHMEAHALTRIL 157  
Db 98 ALSASGVSPSDLSAIAITTIKRGALSLGVGLSFLQVGVOLKPKPIPIHMEAHALTRIL 157

```
Db 142 ALDRAKLEKDSAVAVTIGPGLSLCLRGVKKARKVAANFSLPIYGVHMEHALVARTL 201
Qy 158 T-NKVEPFLVLLISGGCHLALVOGVSDFLLGSLDIAPGMDLKVARRSLIKHPKC 216
Db 202 VEGELSFPMALLISGCHNLVLAKLQGYTOLGTTVDADIGAPDKTKWGLDMH--- 258
Qy 217 STMGGAKEIEHLAKQGNFPHEDIKPLHAKNCDPSFTGLQHYTDKIIMKKEKEGIEKG 276
Db 259 --RSGGPAYEELEAGDAKSVKFNVPKMKYKDCNFSYALKTQVRLAIBAKE----- 308
Qy 277 QILSSAADIATVQHTMACHLVKTRHALFCKQBDLPQNNAVLVASGVASNFYIRPA 336
Db 309 --INRRADIAASFQVAVLHLEKCEKRAIDWALE--LEPSIKHWISGVASNKYVRLR 363
Qy 337 LEITNATQCTLLCPPLCTDNGIMIAMNGIERLAPGLIHDIEGIRYE----- 387
Db 364 LNNIVENKULVCPSPSLCTDNGVMAVMTGJLHFRVG-----RYDPPPRATEPE 413
Qy 388 -----PKCPGLVDISKEVGEA 403
Db 414 DYVYDLRRWPLGEEYAKGRSEA 436
```

## RESULT 5

```
US-10-067-443-25
; Sequence 25, Application US/10067443
```

```
; Patent No. 6642041
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Bristol-Myers Squibb Company
```

```
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
```

```
; FILE REFERENCE: D0073 NP
```

```
; CURRENT APPLICATION NUMBER: US/10/067,443
```

```
; PRIOR FILING DATE: 2002-02-05
```

```
; PRIOR APPLICATION NUMBER: US 60/266,518
```

```
; PRIOR FILING DATE: 2001-02-05
```

```
; PRIOR APPLICATION NUMBER: US 60/282,814
```

```
; NUMBER OF SEQ ID NOS: 71
```

```
; SOFTWARE: PatentIn version 3.0
```

```
; SEQ ID NO 25
```

```
; LENGTH: 179
```

```
; TYPE: PRT
```

```
; ORGANISM: homo sapiens
```

```
US-10-067-443-25
```

```
Query Match 25.8%; Score 549; DB 4; Length 179;
Best Local Similarity 66.5%; Pred. No. 2e-54;
```

```
Matches 119; Conservative 0; Mismatches 0; Indels 60; Gaps 1;
```

```
Qy 38 IVLGIEFSCDDTAAAVDETGNVLGEAIIHSGTEVHLKTGGIYPPAAQQLHRENIQRIYOE 97
Db 1 IVLGIEFSCDDTAAAVDETGNVLGEAIIHSGTEVHLKTGGIYPPAAQQLHRENIQRIYOE 60
Qy 98 ALSASGVSPDLSAIIATTIKPGLALSLGVGSFSQIVGOLKKPRP----- 144
Db 61 ALSASGVSPDLSAIIATTIKPGLALSLGVGSFSQIVGOLKKPRP----- 120
Qy 145 -----IHMEAHALTIIR 156
Db 121 GGAGGCTCANGCCTACTATTAGCTTGACCAATAAGTAGAATTCIHMEAHALTIIR 179

RESULT 6
US-10-067-443-4
; Sequence 4, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
; FILE REFERENCE: D0073 NP
```

```
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-067-443-4
```

```
Query Match 24.7%; Score 524; DB 4; Length 421;
```

```
Best Local Similarity 33.4%; Pred. No. 6.3e-51;
```

```
Matches 130; Conservative 77; Mismatches 146; Indels 36; Gaps 8;
```

```
Qy 39 VLGIEFSCDDTAAAVDETGNVLGEAIIHSGTEVHLKTGGIYPPAAQQLHRENIQRIYOE 98
Db 25 VLGIEFSCDDTAAAVDETGNVLGEAIIHSGTEVHLKTGGIYPPAAQQLHRENIQRIYOE 84
Qy 99 LSASGVSPDLSAIIATTIKPGLALSLGVGSFSQIVGOLKKPRPPIHHMEAHALTIIRLT 158
Db 85 LNDAGTSPKDLDAVAVTVTPGLVIALKEGISAIGFAKGRPLIPVHRRARALSTLLV 144
Qy 159 -NKVEPFLVLLISGGCHLALVOGVSDFLLGSLDIAPGMDLKVARRSLIKHPKCS 217
Db 145 DSVRRFPFSAVNLISGCHALISVADYVKRLYQSVSGSGECIDKVARQGLD-GSEFD 203
Qy 218 TMSGGAKEIEHLAKQGN--RFHEDIKPLHAKNCDPSFTGLQHYTDKIIMKKEKEGIE 274
Db 204 GHVGAAVEILASRABADGHLRYPIFLPVNPKANMFPDQKGSYIMLRLERLKNSESID 263
Qy 275 KQQLSSAADIAATVQHTMACHLVKTRHALFCKQBDLPQNNAVLVASGVASNFYIR 334
Db 264 -----IDPFCASLONTVAHHISSKLIHFFESISECEKLPKO---LVIGGVAAAOYIP 313
Qy 335 RALEITNATQCTLLCPPLCTDNGIMIAMNGIERLAPGLIHDIEGIRYEP----- 388
Db 314 GAISKLSAAHNVTTIVLLSLCTDNAMETAYSQL-----LMLVNSEAIWMPNDIPDT 367
Qy 389 -----KCPGLVDISKEVGEASIKVPOLKM 412
Db 368 IYAHARSDIGTDASSRI-----IDTPRRKL 392
```

## RESULT 7

```
US-10-067-443-28
; Sequence 28, Application US/10067443
```

```
; Patent No. 6642041
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Bristol-Myers Squibb Company
```

```
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
```

```
; FILE REFERENCE: D0073 NP
```

```
; CURRENT APPLICATION NUMBER: US/10/067,443
```

```
; PRIOR FILING DATE: 2002-02-05
```

```
; PRIOR APPLICATION NUMBER: US 60/266,518
```

```
; PRIOR FILING DATE: 2001-02-05
```

```
; PRIOR APPLICATION NUMBER: US 60/282,814
```

```
; NUMBER OF SEQ ID NOS: 71
```

```
; SOFTWARE: PatentIn version 3.0
```

```
; SEQ ID NO 28
```

```
; LENGTH: 421
```

```
; TYPE: PRT
```

```
; ORGANISM: Caenorhabditis elegans
```

```
US-10-067-443-28
```

```
Query Match 24.7%; Score 524; DB 4; Length 421;
Best Local Similarity 33.4%; Pred. No. 6.3e-51;
Matches 130; Conservative 77; Mismatches 146; Indels 36; Gaps 8;
```

```

QY 39 VLGIETSCDDTAAAVVDETVNLGEAISHQTEVHLKTCGIIVPRAAQOLHRENIQRIYVEA 98
DB 25 VLGIETSCDDTAAVAVNEKEKELISERTTERALICQOQGINSPVCAQHRENLPLIIEKC 84
QY 99 LLSAGVSPSDLSAIAATTIKPGIALSLVGLSFSIQLVQOLKKPFIPIHMEAHALTIRLT 158
DB 85 LNDAGTSBKDLDAVAVTVPGLVIALKEGISALIAFKAHRRLPLIPVHMRHALSILLY 144
QY 159 -NKVEFPVLVLLISGGHCLLALVGVSPDLLGKSLDIAPGMDLKVARRLSLIGHPECS 217
DB 145 DDSVRFPPSAVLLSGGHALISVAEDVEFKLYGVGSVSGSPCEIDIKVAQLDGL- GSEPD 203
QY 218 TMSGKAIEHLAKQGN---RHPFDIKPPLHAKNCDPFTGQHTYDKIMKEKEGEGIE 274
DB 204 GIHVGAIVEILASRASADGHLRYFPLENPVKANNFQIKSYIANLERLKRNSETSID 263
QY 275 KGOILSSADIAATVOHTMACHLVKRTIRAILFCQRDILLPQNNAVLVASGVASNFYIR 334
DB 264 -----IPDFCASIQNTVAHHSISKHIFPESLSEQEKLPQ---LVIGGVAAHQYIF 313
QY 335 RALELTNATQCTLLCPPRLCTDNGIMIANNGIRLAKGLIHDIEGIRYEP----- 388
DB 314 GAISKLSAAHVVTIKVLLSLCTDNAEMIAVSGL-----LMLVNRSEAIWVRPNDIDPT 367
QY 389 -----KCPIDVDISKVEGASIKVPOLKM 412
DB 368 IYAHASDIDGTASEI-----IDTPRKXL 392

```

## RESULT 8

```

US-09-540-236-2726
; Sequence 2726, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2726
; LENGTH: 350
; TYPE: PRT
; ORGANISM: M.catarhalis
; US-09-540-236-2726

```

Query Match 23.6%; Score 501; DB 4; Length 350;

Best Local Similarity 34.5%; Pred. No. 2e-48; Indels 44; Gaps 8;

```

DB 126; Conservative 56; Mismatches 139;
QY 39 VLGIETSCDDTAAAVVDETVNLGEAISHQTEVHLKTCGIIVPRAAQOLHRENIQRI 94
DB 6 VLGIETSCDDTAAVAVNEKEKELISERTTERALICQOQGINSPVCAQHRENLPLIIEKC 65
QY 95 VQELASAGVSPSDLSAIAATTIKPGIALSLVGLSFSIQLVQOLKKPFIPIHMEAHALT 154
DB 66 FNEMDQANITKSIDAVAYTKPGILGALMTGALFGRTLAAGIAPVAVGHMEGHLLA 125
QY 155 IRLT--NKVEFPVLVLLISGGHCLLALVGVSPDLLGKSLDIAPGMDLKVARRLSL 212
DB 126 PLLASDDPPSPFVCLLVSGHMTLVRADGVGYQLSGSIDAVGECFDKTKAKMLK-L-P 184
QY 213 HPECSTMGSKAIEHLAKQGNRFHFDIKPPLHAKNCDPFTGLOHTYDKIMKEKEGEG 272
DB 185 YP-----GGPIETKLAKNGNPHAELEPRPMQH-KGLDPSFGSKMAIHNLIKDTNMAOS 237
QY 273 IEKQILSSADIAATVOHTMACHLVKRTIRAILFCQRDILLPQNNAVLVASGVASNFY 332
DB 228 DP-----ATRADIAASFEYAVVDTLVKKCTKALQMTGIRQ-----LVVAGGVASNGT 284
QY 333 IRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAG-----L 375

```

```

DB 285 IRRLTLETLRQIDASVYAPTELCTDNGIAMIAVAGFCRLSKQSDDLAVRCIIPRDMTWL 344
QY 376 GILHD 380
DB 345 GLEVD 349

```

## RESULT 9

```

US-09-252-991A-17372
; Sequence 17372, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: ABRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17372
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-17372

```

Query Match 23.5%; Score 499.5; DB 4; Length 401;

Best Local Similarity 34.4%; Pred. No. 3.8e-48; Indels 29; Gaps 10;

```

DB 13 KPSKRVYEPPLASFNHPTGLFLHKI-----VLGIETSCDDTAAAVVDETVNLGEAISH 66
DB 31 KASDRKTLVWRKADYHARLRICQALRKPMRVLGLETSCDDTAAVAVNEKEKELISERTTERAL 90
QY 67 SQTEVHLKTCGIIVPRAAQOLHRENIQRIYVEA LLSAGVSPSDLSAIAATTIKPGIALSLGV 126
DB 91 SQDHLRVYGVGVPELASSRDHVRMLPLRQVLDESCTPADIDALAYTAPGLVAGALV 150
QY 127 GLSFSIQLVQOLKKPFIPIHMEAHALTIRLTNKK-VEFPVLVLLISGGHCLLALVGVSPD 185
DB 151 GASCAQAMAPAKGVAPVAVGHMEGHLLAPMLEQPPRFVAVLVSGHQLVAVDGIQR 210
QY 186 FULLGKSLDIAPGMDLKVARRLSLIKHPCESTMGSKAIEHLAKQGNRFHFDIKPPLH 245
DB 211 YQLGSSVDDDAAGEAFDXTAKLIGL-GYP-----GGPEIARLAEKGTGPFVFPFRPMTD 263
QY 246 AKKCDPSTFGLOHTYDKIMKEKEGIEKQILSSADIAATVOHTMACHLVKRTIRAI 305
DB 264 RPLGIDPSFSGLKTFPTN-TWQRCVEAGDSEQ---TRCDIALAFQAVAVETLLIKCRAL 319
QY 306 LFCQRDILLPQNNAVLVASGVASNFYIRALELTNATQCTLLCPPRLCTDNGIMIANW 365
DB 320 ---KQYGL-KN--LVLAGGVASANOALISGLEKMGKGVFYARPRCTDNGANIAY 371
QY 366 NGIERLRAGIGLIHDEIGIRYEPKCP 392
DB 372 AGCGRLLAG---QHDPALISVQPRWPM 395

```

## RESULT 10

```

US-08-087-797-3
; Sequence 3, Application US/08087797
; Patent No. 5543312
; GENERAL INFORMATION:
; APPLICANT: Mellora, Alan
; APPLICANT: Lo, Reggie Y.C.
; APPLICANT: Abdullah, Khalid M.
; TITLE OF INVENTION: Pasteurella Haemolytica
; TITLE OF INVENTION: Glycoprotease

```

TITLE OF INVENTION: Gene and the Purified Enzyme  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.  
STREET: 1211 East Morehead Street,  
CITY: Charlotte  
STATE: No. 5543312th Carolina  
COUNTRY: United States  
ZIP: 28234  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087,797  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Layton, Jr., Samuel G.  
REGISTRATION NUMBER: 22807  
REFERENCE/DOCKET NUMBER: 3374-80  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 704 377 1561  
TELEFAX: 704 334 2014  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 342 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-087-797-3

Query Match 23.0%; Score 488; DB 1; Length 342;

Best Local Similarity 35.4%; Pred. No. 66-47; Mismatches 126; Conservative 51; Indels 24; Gaps 8;

39 VLGIETSCDDTAAAVDGTGVNVLGEAIIHSQTEVHLKGTGIVPPAAQOLRENIQRIYQEA 98  
3 VLGIEISCDTETGAIYDDKGLANQIYSGVKLHAYGGVVELASRDHVRKTVPLIQAA 62  
99 LSASGVSPDLSAIIATTKIPGLALSIGVLSFSLQVLVQGLKKPPIPIHMEAHALTIRL 157  
63 LKESGLTAKDIDAVAYTAGPGLVGLVIGATVGRSLAFAMVDPAIPVHMEGHLLAPMLE 122  
158 TNKVEPPLVLLISGCHLLALVQGVSDPLILGKSLDIAPGMDLKVARRLSLIRPECS 217  
123 DNPPEFPFVALLVSGHTQLISVTIGIYELLGESIDDAAGAFDXTAKLGL-DYP--- 178  
218 TMSGKAIEHLAKOGNRFHPDIKPLHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKQ 277  
179 --GGPILSKMAAQGTAGFPFPRPMTDRPGLDFSGSKTFAANTIRDNXXXGXTDDQ 235  
278 ILSSAADIATVQHTMACHLVKTRTALIFCKQDILLPQNN-AVLVASGVASNFYIRRA 336  
236 ---TRADIARAFEDAVVDITMIKCRG-----LDQYKRLVMAGGVANRILRAK 283  
337 LEILTNATQCTLLCPPLCTDNGIMIAMNGIERLAGIGLIHIDIEGIRYERKCP 392  
284 LAEMMKRRGRGVFVARPERFCTDNGAMIAVAGVRFKA--GAVADL-GVSVRRMPL 336

RESULT 11  
US-09-543-681A-6513  
Sequence 6513, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 2709.1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A

CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 6513  
LENGTH: 357  
TYPE: PRT  
ORGANISM: Proteus mirabilis  
US-09-543-681A-6513

Query Match 22.7%; Score 482.5; DB 4; Length 357;

Best Local Similarity 34.2%; Pred. No. 2:6e-46; Mismatches 149; Indels 31; Gaps 8;

39 VLGIETSCDDTAAAVDGTGVNVLGEAIIHSQTEVHLKGTGIVPPAAQOLRENIQRIYQEA 98  
20 VLGIEISCDTETGAIYDDKGLANQIYSGIKHAYGGVVELASRDHVRKTVPLIQAA 79  
99 LSASGVSPDLSAIIATTKIPGLALSIGVLSFSLQVLVQGLKKPPIPIHMEAHALTIRLT 158  
80 LKEANTAKDIDAVAYTAGPGLVGLVIGATVGRSLAFAMVDPAIPVHMEGHLLAPMLE 139  
159 NKV-EPPPLVLLISGCHLLALVQGVSDPLILGKSLDIAPGMDLKVARRLSLIRPECS 217  
140 EKPDPFPVALLVSGHTQLISVTIGIYELLGESIDDAAGAFDXTAKLGL-DYP--- 195  
218 TMSGKAIEHLAKOGNRFHPDIKPLHAKNCDPSFTGLQHTVDKIIMKKEKEGIEKQ 277  
196 --GGPVLKMAQGVGRFVPRPMTDRPGLDFSGSKTFAANTIRDNDS----- 246  
278 ILSSAADIATVQHTMACHLVKTRTALIFCKQDILLPQNN-AVLVASGVASNFYIRRA 336  
247 --QTRADIARAFEDAVVDITMIKCRG-----LEQGFRLVMAGGVANRILRAK 295  
337 LEILTNATQCTLLCPPLCTDNGIMIAMNGIERLAGL-GIHDIEGIRYERKCP 392  
296 MAMIMEQLGGEVYARPELCTDNGAMIALAGMIRFGKTEGPL-----GVTVRRMPL 348

RESULT 12

US-08-087-797-2  
Sequence 2, Application US/08087797  
Patent No. 5543312

GENERAL INFORMATION:  
APPLICANT: Mellors, Alan  
APPLICANT: Lo, Reggie Y.C.  
TITLE OF INVENTION: Abdullah, Khalid M.  
TITLE OF INVENTION: Pasteurella Haemolytica  
TITLE OF INVENTION: Glycoprotease  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.  
STREET: 1211 East Morehead Street,  
CITY: Charlotte  
STATE: No. 5543312th Carolina  
COUNTRY: United States  
ZIP: 28234  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087,797  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Layton, Jr., Samuel G.  
REGISTRATION NUMBER: 22807  
REFERENCE/DOCKET NUMBER: 3374-80  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 704 377 1561



QY 97 EALSASGVSPDLSAIAATTIKRGALSTGVGSLFSLQVLVGQIKKPPPIHMEALTR 156  
DB 89 DALVEAGVSAEDLSAIAATVYGPGLVGLIGISAAPAMAHQPLIPVNHAGHIAAR 148  
QY 157 LTNKVEFPPLVLLISGCHCLALVGVSDPFLIGKSLDIAPGMDLKVARRLSLIKHPBC 216  
DB 149 LVKRFQFPLMLLVSGGHTLVYMOEDGSYEIIGETRDDAAGEAYDKGRVGL----- 202  
QY 217 STMSGKALIEHLAKQGNRFHFDIKPPLHAKNCDPSTGLQHTVDKIMK-KEKEBEGIE 274  
DB 203 -SYPSGKEIDQLAHQGNHYH--PRAMIHEDNYDFPSGLKSAFLINLVHNAQORGEDLD 259  
QY 275 KGQILSSADIAATVQHTMACHLVKTRTALIFCKQBDLLPQNNAV--LVASGVASNFY 332  
DB 260 KN-----DLAASFQASVIDVLITKTLRA---C-----QNYEVKQIVVAGVANAQ 302  
QY 333 IRRALIELTNA--TQCTLLCPPRLCTDNGIMI 363  
DB 303 LREGLOALASAKLPEVELVTPPLRLCGDNAAIMI 335

## RESULT 15

US-08-987-121A-4  
; Sequence 4, Application US/08987121A  
; GENERAL INFORMATION:  
; APPLICANT: Hoskins, Jo Ann  
; APPLICANT: Tang, Joseph Chiou-Chung  
; APPLICANT: Treadway, Patti Jean  
; TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence  
; TITLE OF INVENTION: Gcp  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: U.S.  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/987,121A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Webster, Thomas D.  
; REGISTRATION NUMBER: 39,872  
; REFERENCE/DOCKET NUMBER: X-  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3334  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 336 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-987-121A-4

Query Match 21.7%; Score 461; DB 3; Length 336;

Best Local Similarity 34.5%; Pred. No. 7.3e-44;

Matches 116; Conservative 58; Mismatches 132; Indels 30; Gaps 8;

QY 37 KIVIGIETSCDDTAADVDEGTGVGBAISHQTEVHLKTGGLVPPAAOQLHRENIQRIYQ 96  
DB 4 KYILAFETSCDRTVAIVAKNDELISNVIASQIESHKRFQGVPEVASRHHVEVITACTIE 63  
QY 97 EALSASGVSPDLSAIAATTIKRGALSTGVGSLFSLQVLVGQIKKPPPIHMEALTR 156  
DB 64 EALAEAGITEEDVAVAVTYGPGLVGALLVGLSAKAPAMAHGLPLIPVNHAGHIAAQ 123

QY 157 LTNKVEFPPLVLLISGCHCLALVGVSDPFLIGKSLDIAPGMDLKVARRLSLIKHPBC 216  
DB 124 SVEPLEFPPLIALLVSGGHTLVYVSEAGDYKIVGETRDDAVGEAYDKGRVGL----- 177  
QY 217 STMSGKALIEHLAKQGNRFHFDIKPPLHAKNCDPSTGLQHTVDKIMKKEBEGIE 276  
DB 178 -TYPAGREIDELAHQHDY-YPFRAMIKEDNLEFSGLSKAFINLHNAE-----QKG 230  
QY 277 QILSSADIAATVQHTMACHLVKTRTALIFCKQBDLLPQNNAVLVASGVASNFYIRRA 336  
DB 231 ESLSL-EDLCASFOAAVMDIIMAKTKAL-----EKYPK--TLVAGGVAAANKGIRER 281  
QY 337 LEILTNAQCTLLCPPRLCTDNGIMI-----KN 366  
DB 282 L--ATEITDVNVIIPPLRLCGDNAAIMAAVASEMN 315

Search completed: November 10, 2005, 16:54:34  
Job time : 35.3 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 10, 2005, 16:59:00 ; Search time 1374.75 Seconds  
(without alignments)  
2490.420 Million cell updates/sec

Title: US-10-649-273-2  
Perfect score: 2125  
Sequence: 1 MLILTKTAGVFPPKSKRRVY.....DISKEVGASIVPOLKKEI 414

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 9794790 seqs, 413490567 residues

Total number of hits satisfying chosen parameters: 19589560

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPRO.spool/p/US1064273/runat\_02112005\_091339\_15609/app.query.fasta\_1.1429  
-DB=published.Applcations\_NA -QPM=faetap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=numan40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pcio -NORM=ext -HEARSTIE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US1064273 @CNC 1.1 1034 @runat\_02112005\_091339\_15609  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEBUERT -NEG SCORES=0 -WAIT -DSPELLOC=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09C\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*  
20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq:\*  
21: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq:\*  
22: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq:\*  
23: /cgn2\_6/ptodata/1/pubpna/US10I\_NEW\_PUB.seq:\*  
24: /cgn2\_6/ptodata/1/pubpna/US10I\_NEW\_PUB.seq:\*  
25: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq:\*  
26: /cgn2\_6/ptodata/1/pubpna/US11A\_NEW\_PUB.seq:\*  
27: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
28: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2125	100.0	2197	15	US-10-067-443-1	Sequence 1, Appl1
2	2125	100.0	2197	19	US-10-649-273-1	Sequence 1, Appl1
3	2125	100.0	2197	19	US-10-651-722-1	Sequence 1, Appl1
4	2125	100.0	2572	22	US-10-480-988-36	Sequence 36, Appl1
5	2090.5	98.4	1387	15	US-10-067-443-21	Sequence 21, Appl1
6	2090.5	98.4	1387	19	US-10-649-273-21	Sequence 21, Appl1
7	2090.5	98.4	1387	19	US-10-651-722-21	Sequence 21, Appl1
8	2090.5	98.4	1387	22	US-10-887-553A-1047	Sequence 1047, Ap
9	2088	98.3	1245	15	US-10-012-140-6	Sequence 6, Appl1
10	2088	98.3	1820	15	US-10-012-140-4	Sequence 4, Appl1
11	1944	91.5	2208	18	US-10-094-749-400	Sequence 400, App
12	1944	91.5	2890	21	US-10-723-860-7447	Sequence 7447, App
13	1747	82.2	1416	18	US-10-120-988-177	Sequence 177, App
14	1385	65.2	1526	15	US-10-067-443-23	Sequence 23, Appl1
15	1385	65.2	1526	19	US-10-649-273-23	Sequence 23, Appl1
16	1385	65.2	1526	19	US-10-651-722-23	Sequence 23, Appl1
17	1186.5	55.8	14364	15	US-10-651-722-20	Sequence 20, Appl1
18	1186.5	55.8	14364	19	US-10-649-273-20	Sequence 20, Appl1
19	1186.5	55.8	14364	19	US-10-651-722-20	Sequence 20, Appl1
20	714.5	33.6	1601	26	US-11-097-143-34190	Sequence 34190, A
21	687.5	32.4	3656	26	US-11-097-143-34189	Sequence 34189, A
22	682	32.1	1917	19	US-10-424-599-66417	Sequence 66417, A
23	599	28.2	2734	24	US-10-450-763-20426	Sequence 20426, A
24	578	27.2	1628	20	US-10-437-963-11249	Sequence 11249, A
25	531.5	25.0	4360	15	US-10-081-051-2	Sequence 2, Appl1
26	524	24.7	1146	18	US-10-282-122A-14674	Sequence 14674, A
27	502	23.6	94750	19	US-10-672-787-38	Sequence 26972, A
28	502	23.6	1032	18	US-10-282-122A-31043	Sequence 31043, A
29	494.5	23.3	1026	9	US-09-815-242-7701	Sequence 7701, Ap
30	492.5	23.2	1026	18	US-10-282-122A-30016	Sequence 30016, A
31	492.5	23.2	1026	24	US-10-958-216-49	Sequence 49, Appl1
32	492.5	23.2	1026	24	US-10-958-216-51	Sequence 51, Appl1
33	490.5	23.0	1194	21	US-10-425-115-167432	Sequence 167432, Ap
34	488.5	23.0	1014	9	US-09-815-242-6207	Sequence 6207, Ap
35	484.5	22.8	1014	18	US-10-282-122A-20511	Sequence 20511, A
36	484.5	22.8	1014	18	US-10-282-122A-39301	Sequence 39301, A
37	484.5	22.8	1014	18	US-10-282-122A-41977	Sequence 41977, A
38	483.5	22.8	1014	18	US-10-282-122A-31254	Sequence 31254, A
39	482.5	22.7	1014	9	US-09-815-242-9682	Sequence 9682, Ap
40	481.5	22.4	1029	18	US-10-282-122A-22020	Sequence 22020, A
41	475.5	22.4	1830121	21	US-10-329-670-1	Sequence 1, Appl1
42	475.5	22.4	1830121	18	US-10-156-865-1	Sequence 1, Appl1
43	475.5	22.4	1830121	24	US-10-981-687-1	Sequence 1, Appl1
44	475.5	22.4	1830121	24	US-10-981-687-1	Sequence 1, Appl1
45	475.5	22.4	1830121	24	US-10-981-687-1	Sequence 1, Appl1

## ALIGNMENTS

RESULT 1  
US-10-067-443-1  
; Sequence 1, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; TITLE OF INVENTION: SPINAL CORD, MP-1  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 2197  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (231)..(1472)  
US-10-067-443-1

## Alignment Scores:

Pred. No.:	3,35e-247	Length:	2197
Score:	2125.00	Matches:	414
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-10-649-273-2 (1-414) x US-10-067-443-1 (1-2197)

```
Qy      1 MetLeuIleLeuThrIysThraIagIValIphePheIysProSeIysaIrgIysValIYr 20
Db      231 ATGCTAATCTTGACTAGACAGCTGAGAGATTGTTTAAACCATCAAAAGAAAGTTTAT 290
Qy      21 GIuPheLeuAArgSerPheAsnPheHisProGIYThLeuPheLeuHisIysIleValIleu 40
Db      291 GAATTTTAAAGAAAGTTTAAATTTTCCTCGAAACCTAATTTCTCAATAAAATGATTTG 350
Qy      41 GIYIleGIuThSerCySaspAspThrAlaAlaValIAspGIuThrgIYAsnVal 60
Db      351 GGAAATTGAACCTAGTTGTGATGATACAGACGCTGCTGTGTGATGAAACTGAAATGTG 410
Qy      61 LeuGIYgIuAlaIleHisSerGIuThrgIuValIleuIysThrgIYIleValIPro 80
Db      411 TTGGAGGAAGCAATACATCCCAAACTGAAGTTCATTAAACAGGTGGATTTGCTCT 470
Qy      81 ProAlaIleGIuIleuHisIaIrgIuAsnIleGIuAArgIleValGIuIleuAlaSer 100
Db      471 CCAGACGCTCAACAGCTTCACAGAGAAATATTCACAGATATGACAAAGAGCTCTTCT 530
Qy      101 AlaSerGIYValSerProSerAspIleuSerAlaIleAlaThrIleIysProGIYleu 120
Db      531 GCCAGTGAAGTCTCTCAAGTGAACCTCTTCAGCAATTCGCACTCCATTAACAGGACTT 590
Qy      121 AlaLeuSerLeuGIYValGIYleuSerPheSerLeuGIuIleuValGIYGIuIleuIys 140
Db      591 GCTTTTAAGCTGGAGAGTGGCTTATCATTTAGCTTACAGCTGGTAGGACAGTTAAAAAG 650
Qy      141 PropheIleProIleHisIsmecGIuAlaHisAlaLeuThrIleAArgLeuThrAsnIys 160
Db      651 CCATTCATTCATTCATCATCATGAGGCTCATGCACTTCTATTAGTTCAGCAACATTA 710
Qy      161 ValGIuPheProPheLeuValIleuLeuIleSerGIYHisCySleuLeuAlaLeuVal 180
Db      711 GTAGAAATTTCTTTTATGTTCTTTGATTTCTGGAGGTACTGTCTGTGGGCAATAGTT 770
Qy      181 GIuGIYValSerAspPheLeuLeuLeuGIYIysSerLeuAspIleAlaProGIYAspMet 200
Db      771 CAAAGAGTTTCACATTTTCTGCTTCTTGAAAAGTCTTTTGACATAGACACAGGTGACATG 830
Qy      201 LeuAspIysValAlaIArgAArgLeuSerLeuIleIysHisProGIYCySerThMetSer 220
Db      831 CTTGACAAAGTGGCAAGAAGACTTTCTTTATATAAACATCCAGAGTCTCCACCATAGT 890
Qy      221 GIYGIYIysValIleGIuHisIleuAlaIysGIuIleAsnAArgPheHisIleAspIleIys 240
Db      891 GGTGGGAAAGCCATTAACATTTGGCCAAACAGGAATATGATTTCAATTTTGACATCAA 950
Qy      241 ProProLeuHisIleAlaIysAsnCySaspPheSerPheThrgIYleuGIuHisIleValIThr 260
Db      951 CTTCCCTTGATCATGCTAATAAATGTGATTTTCTTTTACTGACATTCAACACGTTACT 1010
```

```
Qy      261 AspIysIleIleuIysIysGIuIysGIuGIuGIYIleGIuIysGIYGIuIleuSer 280
Db      1011 GATTAATAATTAATGAAAAAGAAAAGAGAAAGTATGAGAAAGGGGCAATCTGCTCT 1070
Qy      281 SerAlaAlaAspIleAlaIleAlaThrValGIuHisIleThrIleAlaCySHIleuValIysArg 300
Db      1071 TCAGCAGCAGACATATGCTGCCACAGTACACACACAAATGCGATTCATCTTGTAAAAAGA 1130
Qy      301 ThrHisArgAlaIleLeuPheCySlyGIuAArgAspLeuLeuProGIuAsnAlaVal 320
Db      1131 ACAATGGGCTATCTGTTTGTGTGAGAGAGACTTGTACTTCATAATTAATGACATA 1190
Qy      321 LeuValAlaSerGIYGIYValAlaIleSerAsnPheIYIleAArgAlaIleuGIuIleu 340
Db      1191 CTGTTGCACTGTGTGTGTGTCGCAAGTAACTTCATATTCGCAAGGCTCTGGAATTTTA 1250
Qy      341 ThrAsnAlaThrGIuIysThrLeuLeuCySProProAArgLeuCySThrAspAsnGIY 360
Db      1251 ACAACGCCAACACAGTGCACCTTGTGTGTCTCTCCACAGATATGACATGATTAATGCGC 1310
Qy      361 IleMetIleAlaIleAArgGIYIleGIuAArgLeuAArgIleuGIYIleLeuHisAsp 380
Db      1311 ATTATGATTCATGCAAGAAATGATTAAGAAAGCTACGTGCTGTGGCATTTTACATGAC 1370
Qy      381 IleGIuGIYIleAArgIYrgIuProIysCySProLeuGIYValAspIleSerIysGIuVal 400
Db      1371 ATGACAGGCACTCCGCTTATGAACCAAAATGTCCTCTTGGAGTAGACATATCAAAAGAACTT 1430
Qy      401 GIYGIuAlaSerIleIysValIProGIuIleuIysIsmecGIuIle 414
Db      1431 GGAGAACTTCCATTAAGTACCAATTAATAATGAGATA 1472
```

## RESULT 2

US-10-649-273-1  
Sequence 1, Application US/10649273  
Publication No. US20040043407A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
FILE REFERENCE: D0073 CNT  
CURRENT APPLICATION NUMBER: US/10/649, 273  
CURRENT FILING DATE: 2003-08-27  
PRIOR APPLICATION NUMBER: US 60/266, 518  
PRIOR FILING DATE: 2001-02-05  
PRIOR FILING DATE: 2001-02-05  
PRIOR FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/282, 814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 1  
LENGTH: 2197  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (231)..(1472)  
US-10-649-273-1

## Alignment Scores:

Pred. No.:	3,35e-247	Length:	2197
Score:	2125.00	Matches:	414
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	19	Gaps:	0

US-10-649-273-2 (1-414) x US-10-649-273-1 (1-2197)

```
Qy      1 MetLeuIleLeuThrIysThraIagIValIphePheIysProSeIysaIrgIysValIYr 20
Db      231 ATGCTAATCTTGACTAGACAGCTGAGAGATTGTTTAAACCATCAAAAGAAAGTTTAT 290
```

QY 21 GIUPELEUARGSERPHEANPHEHISPROGLYTHLEUPHELEUHSILYSIIIEVALLEU 40  
DB 291 GAATTTTAAAGAGTTTAAATTTTCCTCGAAGCACTATTTCTCATATAAATAGATATG 350  
QY 41 GIIYIIIEGLIUTHSERCYSAAPSPTRHRAIAIAIValaValaAspGIUTHRGIVANVAI 60  
DB 351 GGAAATTTGAAACTAGTGTATGATATGACAGAGCTGTGTGTGATGATAAATGGAAATGTG 410  
QY 61 LEUGIYGLIUALIIEHISSESGINTHNGIUVAlHIIELEUYSTRHGIYGLIIEVALPRO 80  
DB 411 TTGGAGAGAGCAATACATTCCTCCAACTGAAGTTCAITTTAAAAACAGTGGGATTGTTCTCT 470  
QY 81 PROAIAIAGIINGIINUHSIARGIUAANIIEGLIARGIIIEVALGINGIUALALEUSER 100  
DB 471 CCAGAGAGCTCAACAGCTTCACAGAGAAATATTCAACGAATAGTACAAAGAAAGCTTTCTTCT 530  
QY 101 AIASESGIYVALSERPROSERAPLEUSERAIIIEAIAThrThrIIElySPROGLYLeu 120  
DB 531 GCCAGTGGAGTCTCTCCAGAGTACCCTCAGCAATTGCAACTAACAAACAGAGACTT 590  
QY 121 AIALEUSERLEUGIYVALGILEUSERPHESELEUGIILEUVALGILEUHSILYS 140  
DB 591 GCTTTAAGCTTGGAGTGGCTTATCATTTAGCTTACAGCTGTAGGACAGTTAAAAAG 650  
QY 141 PROPHETIIEPROIIEHISMETGIUALAHISIALEUThRIIEARGLEUTHRAANLYS 160  
DB 651 CCATTCATTTCCCATTCATATAGAGGCTCAGAGCTCACTACTATTAGTTGACCAATPAA 710  
QY 161 VALGIUPEHROPHLEUVALLEUENIIIESEGIYGLIYHISCYALEUENIAlEUVAL 180  
DB 711 GTAGAAATTTCTTTTAACTTTCTTTGATTTCTGGAGGTCACGTCTGTGGCATTAAGTT 770  
QY 181 GINGIYVALSERASPHELEULEUGIYLYSSERLEUASPIIEAIAProGLIYASPMET 200  
DB 771 CAAGAGATTTCAGATTTTCTGCTCTTGAAGACTTTTGACATACACACAGGTGACATG 830  
QY 201 LEUAPLYSIVAlAIAARGARGLEUSERLEUIIELYHISPROGLUCYSESTRHMETSER 220  
DB 831 CTGGCAAGGTGGCAAGAGACTTTCTTTAAATTAACCAAGAGTCTCCACCAATGAGT 880  
QY 221 GIIYGLIYVALAIIIEGLIUIHISLEUVALYSGINGIYASNARGPHEHISPEHAPRIIELYS 240  
DB 891 GGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATTTTCATTTTGAACATCAA 950  
QY 241 PROPROLEUHIIEHISAIYASNSCYSAAPPHESESPHETHRGILEUGINHISVALITR 260  
DB 951 CCTCCCTTGACATCAGCTAAATAATGTGATTTTCTTTTAACTGACCTTCACACAGTTACT 1010  
QY 261 ASPLYSIIIEIEMETLYSILYSGIUYSGIUGIUGIYIIIEGLIUYSGIYGINIIELEUSER 280  
DB 1011 GATTAATAATATATGAAAGAAAGAAAGAAAGGATTTGAGAGAGGGCAATTCGTGCT 1070  
QY 281 SERAIAIASPPIIEAIAIAThrValGINHISThMETALCYSHISLEUVALIYASARG 300  
DB 1071 TCAGAGAGAGCATTTGCTGCCACAGTACAGCACATAGGCATGTCACTTGTGAAAAGA 1130  
QY 301 THRISHAAGAIIELEUPHETCYSLYSGIINARGAPLEUPROGINANBANAIAVAL 320  
DB 1131 AACACATCCGGCTATCTGTTTGTGTAAGCAGAGACTTGTACTCTCAATAATATAGCAGTA 1190  
QY 321 LEUVALAIASESGIYVALAIAASERANPHEIYRIIEAARGAIIIELEUGIUIIELEU 340  
DB 1191 CTGGTGGATCTGGAGTGTGCGAAGTAACTTCTATATCTCGAGAGCTCTGGAAATTTTAA 1250  
QY 341 THRANAIAATHRGINCYSTRHLEUENUCYSPROPROARGLEUCYSTRHASPANGIY 360  
DB 1251 ACAAAAGCAACACAGTGCATTTGTGTGTCTCTCCCAAGACTAGTAAATATGCG 1310  
QY 361 IIESECTIIEAIAITPANSINGIYIIEGIUARGLEUARGAIIIELEUGIYIIELEUHIASAP 380  
DB 1311 ATTATGATTCATGGAATGTGATTTGAAAGACTACGTCGTGGGCAATTTTAACTATGAC 1370

QY 381 IIEGLIUGIYIIIEARGIYRGIUPEUCYSPROLEUGIYVALAASPIIESETRYGLIUVAl 400  
DB 1371 ATAGAAAGCAATCCGCTATGTGAACCAAAATGTCTCTTGGAGTACATATCAAAAGAACTT 1430  
QY 401 GIIYGLIUALASERIIIELYSVALPROGINLEUHSMETGIUIIE 414  
DB 1431 GGAAAGCTTCATATAAATGATACCAATTAATAATGGAGATA 1472  
RESULT 3  
US-10-651-722-1  
; Sequence 1, Application US/10651722  
; Publication No. US20040048302A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1  
; FILE REFERENCE: D0073 DIV  
; CURRENT APPLICATION NUMBER: US/10/651,722  
; PRIOR FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1  
; LENGTH: 2197  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (231)..(1472)  
US-10-651-722-1  
Alignment Scores:  
Pred. No.: 3,35e-247 Length: 2197  
Score: 2125.00 Matches: 414  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
US-10-649-273-2 (1-414) x US-10-651-722-1 (1-2197)  
QY 1 METLEUUIIELEUTHRILYSTRHRAIAAGIYVALPHEPHELYSPROSETRYSAIGLYSVALITYR 20  
DB 231 ATGCTAATCTTGACTAAGACTGACAGAGTGTGTTTAAACATCAAAAGGAAAGTTTAT 290  
QY 21 GIUPELEUARGSERPHEANPHEHISPROGLYTHRLEUPHELEUHSILYSIIIEVALLEU 40  
DB 291 GAATTTTAAAGAGTTTAAATTTTCCTCGAAGCACTATTTCTCATATAAATAGATATG 350  
QY 41 GIIYIIIEGLIUTHSERCYSAAPSPTRHRAIAIAIValaValaAspGIUTHRGIVANVAI 60  
DB 351 GGAAATTTGAAACTAGTGTATGATATGACAGAGCTGTGTGTGATGATAAATGGAAATGTG 410  
QY 61 LEUGIYGLIUALIIEHISSESGINTHNGIUVAlHIIELEUYSTRHGIYGLIIEVALPRO 80  
DB 411 TTGGAGAGAGCAATACATTCCTCCAACTGAAGTTCAITTTAAAAACAGTGGGATTGTTCTCT 470  
QY 81 PROAIAIAGIINGIINUHSIARGIUAANIIEGLIARGIIIEVALGINGIUALALEUSER 100  
DB 471 CCAGAGAGCTCAACAGCTTCACAGAGAAATATTCAACGAATAGTACAAAGAAAGCTTTCTTCT 530  
QY 101 AIASESGIYVALSERPROSERAPLEUSERAIIIEAIAThrThrIIElySPROGLYLeu 120  
DB 531 GCCAGTGGAGTCTCTCCAGAGTACCCTCAGCAATTGCAACTAACAAACAGAGACTT 590  
QY 121 AIALEUSERLEUGIYVALGILEUSERPHESELEUGIILEUVALGILEUHSILYS 140  
DB 591 GCTTTAAGCTTGGAGTGGCTTATCATTTAGCTTACAGCTGTAGGACAGTTAAAAAG 650

Oy		141	ProphellEProIIehISHmetGluAlahIsalAleuthrIlleaNgLeutRasNve	160
Db		651	CCATTTCATTCCTCATTCATCATAGAGCCTCATGCACCTTACTATTAGGTGACCAATAAA	710
Oy		161	VaIGIuhPePProleuValleuLeuIlleserGIyGIYhisCySleueLaAaleVal	180
Db		711	GTAAGAATTTCTTTTTTAGTTCTTTTGATTTCTGAAGTCACTGTCTGTGGCATTAGTT	770
Oy		181	GIInGIyAlSeRaSPheLeuLeuLeuGIyLySserLeuaspIlleaIproGIyaspMet	200
Db		771	CAAAGAGTTTCAGATTTTCTGCTCTTGAAAAGCTTTTGACAATAGCACCAAGGTGACATG	830
Oy		201	LeuAsplysValAlahagAgLeuSerleuIlleyshISProGIUCysSerThMeSer	220
Db		831	CTTGACAAGGTGGCGAAGAGACTTTCTTTATTAATAAACATCCAGGTCTCCACCATAGT	890
Oy		221	GIyGIyLysSaIalleGIunHsleuAlalySGInGIyASnArgPhenIspeAspIlleys	240
Db		891	GGTGGGAAGCCATAGAACATTGTGCCAAACAGGAANTAGATTTCATTTTGACATCAA	950
Oy		241	ProProleuHSHIsalalySaSnCySaSPheSerPheThrgIyleuGIunHSvalThr	260
Db		951	CCTCCCTTGACATCATGCTAAAAATTGTGATTTTCTTTTACTGACCTTCAACAGTTACT	1010
Oy		261	AspLySIlellemeLySLySGlUlySGlUGlUGlYIleGIuLySGlYgInlleuSer	280
Db		1011	GATAAATATTAATAAGAAAAGAAAAGAGAGAGATATTTGGAAAGGGGCAAAATCTGTCT	1070
Oy		281	SerIlaIlaapIlealaIaThVaGIunHsThrMeAlaCyShIsleuVallysArg	300
Db		1071	TCAAGAGACATGCTGCGCACAGTACAGCACACATGGCATGTCACTTTGTAAANAGA	1130
Oy		301	ThrlhsarGaIalleLeuPheCysLySGlnArgspLeuLeuProGIUnasnAnaIaVal	320
Db		1131	ACACATCGGGCTATTCTGTTTGTAGACAGAGAGACTGTTCACCCAATAATAGCAGTA	1190
Oy		321	LeuValAlaSerGIyGIyValAlaSeranPheTyrlleaAgAgAlaleuGIunlleu	340
Db		1191	CTGGTTGCATCTGGGTGGTGGCGCAAGTAACCTTATATCCGACAGGCTCTGGAAAATTTTA	1250
Oy		341	ThrsnaIaThngInCysThreuleucCysPropoArGleuCysThrAspAnGIy	360
Db		1251	ACAAACGACACACAGTGCATTTGTTGTCTCTCCACAGCTRAGCATGATTAATGCG	1310
Oy		361	IlemecilleaIatyrAnGIyIleGIuArgLeuArgAlaGIyleuGIyIleuHIsAp	380
Db		1311	ATTATGATTCATGCATGAATGATTAATAAGACTAGCTGTGGCTTGGCATTTTACATGAC	1370
Oy		381	IleGIuGIyIleArgTyrgIuProLyCySPoleuGIyValaspIleSerLySGlVal	400
Db		1371	ATGAAAGGCATCCGCTATGAACCAAAAAGTCTCTTGAGATGACATATCAAAAGAAAGTT	1430
Oy		401	GIyGIuAlaSerIleIeVAlProGIUnleuIySmetGIulle	414
Db		1431	GGAGAAAGCTCCATAAAGTACCACAAATTAATAAATAGAGATA	1472
RESULT 4				
US-10-480-988-36				
Sequence 36, Application US/10480898				
Publication No. US20050069877A1				
GENERAL INFORMATION:				
APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;				
APPLICANT: SWARNAKAR, Anita; HARILAL, April J.A.;				
APPLICANT: TRAN, Bao; DUGAN, Brendan M.;				
APPLICANT: WARREN, Bridget A.; ISON, Craig H.;				
APPLICANT: HONGBELU, Cynthia D.; NGUYEN, Daniel B.;				
APPLICANT: LU, DYUNG AINA M.; LEE, Ernestine A.;				
APPLICANT: YUE, Henry; FORSYTHE, Ian J.;				
APPLICANT: BARROSO, Ines; RAMQUAR, Jayalakmi;				
APPLICANT: GRIFIN, Jennifer A.; LI, Joana X.;				
APPLICANT: YANG, Tuming; THANHAVEIU, Kavitha;				
APPLICANT: GIETZEN, Kimberly J.; DING, Li;				
APPLICANT: BAUGHN, Mariah R.; BOROWSKI, Mark L.;				

APPLICANT: YAO, Monique G.; CHAWLA, Narinder K.;  
APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal,  
APPLICANT: LEE, Sally; BECHA, Shanya D.;  
APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;  
APPLICANT: ELLIOTT, Vicki S.; LOO, Wen;  
APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;  
APPLICANT: LU, Yan; ZEBARJADIAN, Vaganeh  
TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES  
FILE REFERENCE: PF-1040 USN  
CURRENT APPLICATION NUMBER: US/10/480, 988  
CURRENT FILING DATE: 2003-12-16  
PRIOR APPLICATION NUMBER: PCT/US02/19360  
PRIOR FILING DATE: 2002-06-18  
PRIOR APPLICATION NUMBER: US 60/300, 508  
PRIOR FILING DATE: 2001-06-22  
PRIOR APPLICATION NUMBER: US 60/303, 445  
PRIOR FILING DATE: 2001-07-06  
PRIOR APPLICATION NUMBER: US 60/305, 405  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/311, 442  
PRIOR FILING DATE: 2001-08-09  
PRIOR APPLICATION NUMBER: US 60/314, 821  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/315, 992  
PRIOR FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: US 60/378, 205  
PRIOR FILING DATE: 2002-05-03  
NUMBER OF SEQ ID NOS: 56  
SOFTWARE: PERL Program  
SEQ ID NO 36  
LENGTH: 2572  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No: 7632424CBI  
US-10-480-988-36

Alignment Scores:  
Pred. No.: 4.3e-247 Length: 2572  
Score: 2125.00 Matches: 414  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 22 Gaps: 0

US-10-649-273-2 (1-414) x US-10-480-988-36 (1-2572)

QY 1 MetLeuIleuThrIysThrAlaGlyValPhePheLysProSerIysArgIysValTyr 20  
Db 144 ATGCTAACTTGACTGAAGCTGAGGAGTTTATTAAACATCAAAAAGAAAGTTTAT 20  
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIysIleValLeu 40  
Db 204 GAATTTTAAAGAAATTTTAAATTTTCAATCCGGAACATATTCTTCATAAATAGATATG 26  
QY 41 GlyIleGluThrSerCysAspArgThrAlaAlaValAlaIspGluThrGlyAsnVal 60  
Db 264 GGAATTGAAACTAGTGTGATGATACAGACAGCTGCTGTGTGATGAAACTGAAATGTG 32  
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisIleuLysThrGlyGlyIleValPro 80  
Db 324 TTGGAGAGAAATPACATCTCCCAAACTGAAAGTCATTAAACAGAGTGGATTGCTTCT 38  
QY 81 ProAlaIleGlnGlnIleuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 10  
Db 384 CCAGAGCTCAACAGCTTTCACAGAGAAATATTCACAGATATATACAGAGACTTTTCT 44  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 12  
Db 444 GCCATGTGAGTCTCTCCAAAGTACCTCTCAGCAATTGCACATCAATAAAACAGAGACTT 50  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuIleuValGlyIleuLysValys 14



Db 804 GATAAATATATATGAAAAAGAGAGAGTATATTCTAATTAGTAAGTTGAA 863  
Qy 274 -----GluIys 275  
Db 864 CAGATAAATATTTCTGATTTGCTCTAAATAATAGCTGCTATTTCGACAGTATGAGAG 923  
Qy 276 GtGtlnlleuSerSerAlaIaAspIleAlaIaThrValGlnHisThrMetAlaCys 295  
Db 924 GGGGAAATCCGTCTTCAGCAGACAGATGCTGCGACAGTACAGACCAATGGCATGT 983  
Qy 296 HisleuValIysArgThrHisArgAlaIleuPheCysIysGlnArgAspIleuPro 315  
Db 984 CATCTTGAAAAAGAACATCGGCTATTCTGTTTGTGAACAGAGACTTTGTTACCT 1043  
Qy 316 GlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 335  
Db 1044 CAATAATATGCGATCTGGTTCATCTGGTGTGTCGCAAGTAACCTTATATCCGACAG 1103  
Qy 336 AlaIeuGluIleuThrAsnAlaThrGlnCysGthrIleuLeuCysProProArgIleu 355  
Db 1104 GCTCGAAATTTTAAACAACCAACAGTGCATCTTGTGTCTCTCCAGACATA 1163  
Qy 356 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgIleuArgAlaGlyLeu 375  
Db 1164 TGCACCTGATATGCGATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1223  
Qy 376 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProIysCysProIeuGlyValAsp 395  
Db 1224 GGCATTTTACATGACATTAAGGCAATCCGCTATGAACCAAAATGCTCTTGGAGTAGAC 1283  
Qy 396 IleSerIysGluValGlyGluAlaSerIleIysValProGlnIleuIysMetGluIle 414  
Db 1284 ATATCAAAAAGAGTTGGAGAGGCTTCATATAAAGTACCAATTAATAAATGAGAGATA 1340

RESULT 6  
US-10-649-273-21  
; Sequence 21, Application US/10649273  
; Publication No. US20040043407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEINASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649, 273  
; CURRENT FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 21  
; LENGTH: 1387  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-649-273-21

Alignment Scores:  
Pred. No.: 2,57e-243 Length: 1387  
Score: 2090.50 Matches: 412  
Percent Similarity: 93.85% Conservative: 0  
Best Local Similarity: 93.85% Mismatches: 2  
Query Match: 98.38% Indels: 25  
DB: 19 Gaps: 1

US-10-649-273-2 (1-414) x US-10-649-273-21 (1-1387)

Qy 1 MetLeuIleuPheThrIysThrAlaGlyValPhePheIysProSerIysArgIysValIyr 20  
Db 24 ATGCTATATCTTGCTAAAGACTGACGAGAGTATTTTAAACATCAAAAAGAAAGTTAT 83  
Qy 21 GluPheIeuArgSerPheAsnPheHisProGlyThrIleuPheIleuHisIysIleValIleu 40

Db 84 GATTTTAAAGAGTTTAAATTTTCATCCGAAACACTATTTCTTCAATAAATAGTATG 143  
Qy 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60  
Db 144 GGAATTTGAAACCTAGTGTATGATGATACAGACGCTGCTGTGTGATGATGAACTGAAATGTG 203  
Qy 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisIleuIysThrGlyIleValPro 80  
Db 204 TTGGAGAAAGCAATACATTCCTCAACTGAAGTTCAATTAATAACAGGTGGATGTTGCTCT 263  
Qy 81 ProAlaAlaGlnGlnIleuHisArgGluAsnIleGlnArgIleValGlnGluAlaIleuSer 100  
Db 264 CCAAGAGCTCAACAGCTTCAACAGAGAAATATTCACAGATATGACAAAGAAAGCTCTTCT 323  
Qy 101 AlaSerGlyValSerProSerAspIleuSerAlaIleAlaThrThrIleIysProGlyLeu 120  
Db 324 GCGAGTGAAGTCTCTCAAGTGAACCTTCAGCAATTCGAATCAATTAATAACAGAGACT 383  
Qy 121 AlaIeuSerIeuGlyValGlyIleuSerPheSerIeuGlnIleuValGlyIleuIysIys 140  
Db 384 GCTTTAAGCTGGAGAGTGGCTTATCATTTTACCTTACAGCTGGTATGAGACAGTTAAAAAG 443  
Qy 141 ProPheIleProIleHisIleMetGluAlaHisAlaIleuThrIleArgIleuThrAsnIys 160  
Db 444 CCATTCATTCCTCATCATCATATGAGAGCTCATGACATTCATTAAGTTGACCAATATAA 503  
Qy 161 ValGluPheProPheIleuValIleuIleIleSerGlyGlyHisCysIleuIleuAlaIleuVal 180  
Db 504 GTAGAAATTCCTTTTATGATCTTTTGAATTCGAGAGTCACTGTCTGTGGCTTATAGTT 563  
Qy 181 GlnGlyValSerAspPheIleuLeuIeuGlyIysSerIeuAspIleAlaProGlyAspMet 200  
Db 564 CAAGAGATTTCAGATTTTCTGCTTTCGAAAGCTTTGAGACATGACACAGTGCATG 623  
Qy 201 IeuAspIysValAlaArgArgIeuSerIeuIleIysHisProGluCysSerThrMetSer 220  
Db 624 CTTCAGCAAGCTGGCAAAAGCTTTCTTATAATAAATCATCAGAGTCTCCACATGAGT 683  
Qy 221 GlyIlyValAlaIleGluHisIleuAlaIysGlnIysAsnArgPheHisPheAspIleIys 240  
Db 684 GGTGGAAAGCCATGAGCATTTGGCCAAACAGAAATGATTTCAATTTGACATCAAA 743  
Qy 241 ProProIeuHisHisAlaIysAsnCysAspPheSerPheThrGlyIleuGlnHisValThr 260  
Db 744 CTTCCCTTGATCATGCTAAATAATGATTTTCTTTACTGACATTCACACAGCTTACT 803  
Qy 261 AspIysIleIleMetIysGlyIysGluIeuGlyIle----- 273  
Db 804 GATTAATATATGAAAAAGAAAGAGAGATATTTCTAATTAGTAAGTTGAA 863  
Qy 274 -----GluIys 275  
Db 864 CAGATAAATATTTCTGATTTGCTCTAAATAATAGCTGCTATTTCGACAGTATGAGAG 923  
Qy 276 GtGtlnlleuSerSerAlaIaAspIleAlaIaThrValGlnHisThrMetAlaCys 295  
Db 924 GGGGAAATCCGTCTTCAGCAGACAGATGCTGCGACAGTACAGACCAATGGCATGT 983  
Qy 296 HisleuValIysArgThrHisArgAlaIleuPheCysIysGlnArgAspIleuPro 315  
Db 984 CATCTTGAAAAAGAACATCGGCTATTCTGTTTGTGAACAGAGACTTTGTTACCT 1043  
Qy 316 GlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 335  
Db 1044 CAATAATATGCGATCTGGTTCATCTGGTGTGTCGCAAGTAACCTTATATCCGACAG 1103  
Qy 336 AlaIeuGluIleuThrAsnAlaThrGlnCysGthrIleuLeuCysProProArgIleu 355  
Db 1104 GCTCGAAATTTTAAACAACCAACAGTGCATCTTGTGTCTCTCCAGACATA 1163  
Qy 356 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgIleuArgAlaGlyLeu 375







NUMBER OF SEQ ID NOS: 1208  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 1047  
 LENGTH: 1387  
 TYPE: DNA  
 ORGANISM: human  
 US-10-649-273-2 (1-414) x US-10-649-273-2 (1-1245)

## Alignment Scores:

Pred. No.: 2.57e-243 Length: 1387  
 Score: 2090.50 Matches: 412  
 Percent Similarity: 93.85% Conservative: 2  
 Best Local Similarity: 93.85% Mismatches: 0  
 Query Match: 98.38% Indels: 25  
 DB: Gaps: 1

US-10-649-273-2 (1-414) x US-10-649-273-2 (1-1245)

QY 1 MetLeuIleuThrThraAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
 Db 24 ATGCTAAATCTTACCTAGAGCTGACAGAGATTTTAAACATCAAAAAGAAAGTTTAT 83  
 QY 21 GluPheLysSerPheAspPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
 Db 84 GAATTTTAAAGAGTTTAAATTTTCACTCGAAACCTATTCTTCAATAAATAGTATG 143  
 QY 41 GlyIleGlyThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60  
 Db 144 GGAATTTGAAACCTAGTTGATGATACAGCAGCTGCTGTGATGATAACTGGAATGTG 203  
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80  
 Db 204 TTGGAGAAAGCAATCAATCCCAACTGAAGTTCAATTAACAGGTGGATTTGCT 263  
 QY 81 ProAlaAlaGlnIleuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 Db 264 CAGCAGCTCAACAGCTTACAGAGAAATATTCACGAAATGATACAGAGAGCTTTCT 323  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
 Db 324 GCCAGTGGAGTCTCTCAAGTGAACCTCAGCAATGCAATCAATAAACAGAGACTT 383  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysVal 140  
 Db 384 GCTTTAAGCTGGAGTGGCTTATCATTTAGCTTACAGCTGGTAGACAGTTAAAG 443  
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
 Db 444 CCAATTCATCCATTCATCATATGAGGCTCATGCACTTATTTAGGTTGACCAATAA 503  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 Db 504 GATGATTTCTTTTATGATTTCTTTGATTTCTGGAGGTCACTGTTGGCATGAT 563  
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
 Db 564 CAAGGAGTTTCAGATTTTCTGCTTCTTGAAGCTTTGACATAGCACAGGTGACATG 623  
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220  
 Db 624 CTTCAGCAGGTGGCAAGAGACTTTCTTTAATTAACATCAAGAGTCTCCACATGAGT 683  
 QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
 Db 684 GGTGGAGAAACCTATAGCATTTGGCCAAACAGAAATATGATTTCAATTTGACATCAA 743  
 QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
 Db 744 CTTCCCTTGATCATGCTAAATAATGATTTTCTTTTCTGACATTCACACAGTACT 803  
 QY 261 AspLysIleIleLeuLysLysGlyLysGlnGlyIle----- 273  
 Db 804 GATTAATAATATATGAAAAAGAAAAAGAGAAAGTATATTTCTAATTAAGTAAAGTTGAA 863

QY 274 -----GluLys 275  
 Db 864 CAGATTAATATTTCTGATTTGTGCTTAAATAAGCTGCTCAATTTGACAGTATGAGAG 923  
 QY 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 295  
 Db 924 GGGCAATCTGCTGCTTACAGAGAGCATTTGCTGCGACAGTACAGACCAATAGGATGT 983  
 QY 296 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLeuGlnArgAspLeuPro 315  
 Db 984 CATCTGTGAAGAAACACATCGGCTTATTTCTTTTGTAGAGAGAGCTTTGATACCT 1043  
 QY 316 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 335  
 Db 1044 CAATAAATAGCAGTATCGTTGATCTGGTGGTGGCAAGTAACTTCTATATCCGAGA 1103  
 QY 336 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProCysArgLeu 355  
 Db 1104 GCTCTGAAATTTTAAACAAACGACACAGTGCATTTGTGTCTCTCCACAGCTA 1163  
 QY 356 CysThrAspAsnGlyIleMetIleAlaTPAsnGlyIleGluArgLeuArgAlaGlyLeu 375  
 Db 1164 TGCACGTAAATGCAATTAATGATGATGAGATGATGATTAAGACCTACCTGCTGCT 1223  
 QY 376 GlyIleLeuHisAspIleGlnGlyIleArgGlyArgProLysCysProLeuGlyValAsp 395  
 Db 1224 GGCATTTTACATACATAGAGGCAATCCGCTAAGAACCAAAAGTCTCTTGAGATGAC 1283  
 QY 396 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
 Db 1284 ATATCAAAAGAGTGGAGAACTTCATTAAGTACCAATTAATAAATGAGATATA 1340

## RESULT 9

US-10-012-140-6  
 ; Sequence 6, Application US/10012140  
 ; Publication No. US2003009017A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Leibny, Kevin R.  
 ; APPLICANT: Kapeller-Libermann, Rosana  
 ; APPLICANT: Glucksmann, Maria A.  
 ; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND  
 ; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES  
 ; TITLE OF INVENTION: THESEOF  
 ; FILE REFERENCE: 381552004900  
 ; CURRENT APPLICATION NUMBER: US/10/012,140  
 ; CURRENT FILING DATE: 2001-11-08  
 ; PRIOR APPLICATION NUMBER: 60/246,768  
 ; PRIOR FILING DATE: 2000-11-08  
 ; PRIOR APPLICATION NUMBER: 60/246,772  
 ; PRIOR FILING DATE: 2000-11-08  
 ; PRIOR APPLICATION NUMBER: 60/249,185  
 ; PRIOR FILING DATE: 2000-11-15  
 ; NUMBER OF SEQ ID NOS: 49  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 6  
 ; LENGTH: 1245  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-012-140-6

## Alignment Scores:

Pred. No.: 4.36e-243 Length: 1245  
 Score: 2088.00 Matches: 407  
 Percent Similarity: 99.03% Conservative: 3  
 Best Local Similarity: 98.31% Mismatches: 4  
 Query Match: 98.26% Indels: 0  
 DB: Gaps: 0

US-10-649-273-2 (1-414) x US-10-012-140-6 (1-1245)

QY 1 MetLeuIleuThrThraAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
 |||||||

Db 1 ATGCTATCTTGACCTAAGACCTGACGAGTCTTTTAAACATCAAAAAGAAAGTTAT 60  
 QY 21 GIUphELeuAArgSerPheAsnPhEhiSPROGlyThrLeuPhELeuHisiLyIleVal1Leu 40  
 Db 61 GAATTTTAAAGAAAGTTTAAATTTTCACTCGAAGACATAATTTCTCATATAAATAGATTC 120  
 QY 41 GIIYIleGIuThrSerCyAspAspThrAlaAlaValAlaAspGIuThrGIyaEnVal 60  
 Db 121 GGAAATGAAACTGATGTGATGATACAGACGCTGCTGTGTGATGATAAACTGAAAATGTG 180  
 QY 61 LeuGIyIuAlaIleHisiSerGIuThrGIuValHisiLeuLyEThrGIyIleVal1Pro 80  
 Db 181 TTGGAGAAAGCAATACATTCCTCAAGTGAAGTTCATTATAAAAAGGTGGATTTCTCT 240  
 QY 81 ProAlaAlaGIuThrLeuHisiArgIuAsnIleGIuArgIleVal1GIuThrLeuSer 100  
 Db 241 CCACAGAGCTCAACAGCTTCAAGAGAAATATTCAACGAATAGTACAGAGAGCTTTTCT 300  
 QY 101 AlaSerGIyValSerProSerAspLeuSerAlaIleAlaThrThrIleLySPROGIyLeu 120  
 Db 301 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGAAATTCGAACTAACCTAAACAGAGACTT 360  
 QY 121 AlaLeuSerLeuGIyValGIyLeuSerPheSerLeuGIuThrValGIyIleuLySIVS 140  
 Db 361 GCTTTAAGCTGGAGTGGAGTGGCTTATCATTTAGCTTACAGCTGGTAGACAGTTAAAAAG 420  
 QY 141 ProPheIleProIleHisiSweGIuAlaHisiAlaLeuThrIleArgLeuThrAsnLyS 160  
 Db 421 CCATTTCATTCCTCATTCATATGAGAGGCTCATGCACTTACTATTAGTTAGTACCAATAA 480  
 QY 161 ValGIuPheProPheLeuVal1LeuLeuIleSerGIyIleHisiCysLeuVal1LeuVal 180  
 Db 481 GTAGAAATTCCTTTTGTAGTTCTTTGTGATTTCTGAGAGGTCACGTGCTGTGGCATTAATT 540  
 QY 181 GIuGIyVal1SerAspPheLeuLeuGIyLySLeuSerLeuAspIleAlaProGIyAspMet 200  
 Db 541 CAAGAGATTTCAGATTTCTGCTTTTGAAGAGTCTTTGGACATAGCAACAGTGAACAG 600  
 QY 201 LeuAspLySValAlaArgArgLeuSerLeuIleLySHisSPROGIyCysSerThrMetSer 220  
 Db 601 CTTGACAAAGTGGCAGAGAGACTTTCTTAATAAATCATCCAGAGTGCCTCCACCATGAGT 660  
 QY 221 GIyGIyLyValAlaIleGIuHisiLeuAlaLySGInGIyAsnArgPheHisiPheAspIleLyS 240  
 Db 661 GGTGGAAAGCCATAGAACATTGGCCAAACAGAAATAGATTTCATTTTACATCAACAA 720  
 QY 241 ProProLeuHisiSHisAlaLyAsnCyAspPheSerPheThrGIyLeuGIuHisiVal1Thr 260  
 Db 721 CTTCCCTTGCATCATGTCTAAATATGCTATTTTCTTTTACCTGGAACCTCAACACGTTACT 780  
 QY 261 AspLySIIleIleMetLyELeuGIuLyGIuGIyIleGIuLyGIyGIuIleLeuSer 280  
 Db 781 GATTAATAATTAAGAAACAGAAACAGAAAGGATTAAGAAAGGCGCAAAATCCGTCT 840  
 QY 281 SerAlaAlaAspIleAlaAlaThrValGIuHisiThrMetAlaCysHisiLeuVal1LySArg 300  
 Db 841 TCAGCAGACAGCATTTGCTGCCACAGTACAGACACAAATGCGATGTCATCTTGTGAAAAGA 900  
 QY 301 ThrHisiArgAlaIleLeuPheCysLySGInArgAspLeuLeuProGIuHisiAsnAlaVal 320  
 Db 901 ACACATCGGGCTATTCGTCTTTGTGTAAGCAGAGACTGTTAACCTCAAAATATATCAGTA 960  
 QY 321 LeuVal1AlaSerGIyGIyValAlaSerAsnPheThrIleArgArgAlaLeuGIuIleLeu 340  
 Db 961 CTGCTTCATCTGAGTGTGTGCGCAAGTACTTATATCCGACAGCTCTGCAAAATTTTA 1020  
 QY 341 ThrAsnAlaThrGIuCysThrLeuLeuCysProProArgLeuCysThrAspAsnGIy 360  
 Db 1021 ACAACGCAACACAGTGCATTTGTGTGTCTCTCCACAGCTATAGCATATATATGCT 1080  
 QY 361 IleMetIleAlaATPAsnGIyIleGIuArgLeuArgAlaGIyLeuGIyIleLeuHisiAsp 380  
 Db 1081 ATATATGATTCATGGAATGATTAAGAAAGACTAGCTGCTGGCATTTTATCATGAC 1140

QY 381 IIEGIuGIyIleArgIyArgIuProLyCyAspProLeuGIyValAspIleSerLyGIuVal 400  
 Db 1141 ATAGAAGCATCCGCTATGAAACCAAAATGCTCTTGGAGTAGACATATCAAAAGAGTT 1200  
 QY 401 GIyGIuAlaSerIleLySVal1ProGIuThrLeuHisiMetGIuIle 414  
 Db 1201 GGAAAGCTTCATTAAGATACCAAGATTAATAAATGAGATA 1242  
 RESULT 10  
 US-10-012-140-4  
 ; Sequence 4, Application US/10012140  
 ; Publication No. US2003009017A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Leiby, Kevin R.  
 ; APPLICANT: Kapeller-Libermann, Rosana  
 ; APPLICANT: Glucksmann, Maria A.  
 ; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81586, AND  
 ; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES  
 ; FILE REFERENCE: 381552004900  
 ; CURRENT APPLICATION NUMBER: US/10/012,140  
 ; PRIOR FILING DATE: 2001-11-08  
 ; PRIOR APPLICATION NUMBER: 60/246,768  
 ; PRIOR FILING DATE: 2000-11-08  
 ; PRIOR APPLICATION NUMBER: 60/246,772  
 ; PRIOR FILING DATE: 2000-11-08  
 ; PRIOR APPLICATION NUMBER: 60/249,185  
 ; PRIOR FILING DATE: 2000-11-15  
 ; NUMBER OF SEQ ID NOS: 49  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1820  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (146) ... (1390)  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1) ... (1820)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-10-012-140-4  
 Alignment Scores:  
 Pred. No.: 7,95e-243 Length: 1820  
 Score: 2088.00 Matches: 407  
 Percent Similarity: 99.03% Conservative: 3  
 Best Local Similarity: 98.31% Mismatches: 4  
 Query Match: 98.26% Indels: 0  
 DB: 15 Gaps: 0  
 US-10-649-273-2 (1-414) x US-10-012-140-4 (1-1820)  
 QY 1 MetLeuIleLeuThrIleThrAlaGIyValPhePheLySPROSerLySArgLyVal1Thr 20  
 Db 146 ATGCTAATCTTGACTAGACTGACAGAGTTTATTTTAAACATCAAAAAGAAAGTTAT 205  
 QY 21 GIUphELeuAArgSerPheAsnPhEhiSPROGIyThrLeuPhELeuHisiLyIleVal1Leu 40  
 Db 206 GAATTTTAAAGAAAGTTTAAATTTTCACTCGAAGACATAATTTCTCATATAAATAGATTC 265  
 QY 41 GIIYIleGIuThrSerCyAspAspThrAlaAlaValAlaAspGIuThrGIyaEnVal 60  
 Db 266 GGAAATGAAACTGATGTGATGATACAGACGCTGCTGTGTGATGATAAACTGAAAATGTG 325  
 QY 61 LeuGIyIuAlaIleHisiSerGIuThrGIuValHisiLeuLyEThrGIyIleVal1Pro 80  
 Db 326 TTGGAGAAAGCAATACATTCCTCAAGTGAAGTTCATTATAAAAAGGTGGATTTCTCT 385  
 QY 81 ProAlaAlaGIuThrLeuHisiArgIuAsnIleGIuArgIleVal1GIuThrLeuSer 100  
 Db 386 CCAGAGCTCAACAGCTTCAAGAGAAATATTCAACGAATAGTACAGAGAGCTTTTCT 445

```

QY 101 AlAserGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLeu 120
Db 446 GCCAGTGAAGTCTCTCCAAAGGACCTTCAGCAATTCGAACTACCAATAAACAAGACACTT 505
QY 121 AlaleuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
Db 506 GCTTTAAGCCTGGAGAGTGGCTTATCACTTACCTTACAGCTGTGTGAGCAAGTAAAAAG 565
QY 141 ProPheIleProIleHisIleMetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 566 CCATTCAATCCCATTCATCATATGAGGCTCATGACATTACTATTAAGTTGACCAATAAA 625
QY 161 ValGlnPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db 626 GTAAATTTCTTTTATTTAGTTCTTTTATTTGATTTCTGAGGCTGCTGTGTGACATTAAGT 685
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
Db 686 CAAGGAGTTTCAATTTCTGCTTTTGGAAAGCTTTTGAACATGACCAAGTGAATG 745
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlnCysSerThrMetSer 220
Db 746 CTTGACAAGTGGCAAGAAAGACTTTCTTTAATAAACAATCCAGAGTGTCCACCAATGAGT 805
QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
Db 806 GGATGGAAAGCCATGGAACATTTGGCCAAACAAAGAAATGATTTGATTTGACATCAAA 865
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 866 CCTCCCTTGACATGCTAAATAATGTGATTTTCTTTTACTGAGACTTCAACAGTTACT 925
QY 261 AspLysIleIleMetLysLysGlnLysGlnGlyIleGlnLysGlyGlnIleLeuSer 280
Db 926 GATTAATAATTAATGAACAAGAAACAAGGAAAGATTTGAGAAGGGCAAAATCTGTCT 985
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
Db 986 TCAGCAGCAGACATGTGCTGCCACAGTACAGCACACATATGTCATCTTGTGAAAGA 1045
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAlaVal 320
Db 1046 ACAACATCGGGCTAATCTGTGTGTGTAAGCAGAGACTGTGTACTCAAAATTAATGCACTA 1105
QY 321 LeuValAlaSerGlyValAlaAspAsnPheThrIleArgArgAlaLeuGlnIleLeu 340
Db 1106 CTGGTTGCACTTGGTGGTGTCCGAGTAACTTCTATATCCGAGAGCTCTGGAATTTTA 1165
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360
Db 1166 ACAACGCAACACAGTGCACCTTTGTGTGTCTCCCTCCAGACTATGCACTGATTAATGGC 1225
QY 361 IleMetIleAlaIleAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db 1226 ATTAATGATTTGACATGATGATTAAGAAAGCTAGCGTGGCTTTGGCAATTTTACATGAC 1285
QY 381 IleGlnGlyIleArgTyrGlnProLysCysProLeuGlyValAspIleSerLysGlnVal 400
Db 1286 ATAGAAAGGATCCGCTATGAACCAAAATGCTCTTGAGATGACATTAACAAAGAAATT 1345
QY 401 GlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 414
Db 1346 GGAGAAAGCTTCCATAAAGTACCAACATTAATAATGAGATA 1387

```

```

/ APPLICANT: WAKAMATSU, AI
/ APPLICANT: SATO, HIROYUKI
/ APPLICANT: ISHII, SHIZUKO
/ APPLICANT: YAMAMOTO, JUN-ICHI
/ APPLICANT: ISONO, YUUKO
/ APPLICANT: HIO, YUKI
/ APPLICANT: OTSUKA, KAORU
/ APPLICANT: NAGAI, KEIICHI
/ APPLICANT: IRIE, RYOTARO
/ APPLICANT: TAMECHIKA, ICHIRO
/ APPLICANT: SEKI, NAOHICO
/ APPLICANT: YOSHIKAWA, TSUTOMU
/ APPLICANT: OTSUKA, MOTOYUKI
/ APPLICANT: NAGAHARI, KENJI
/ APPLICANT: MASUHO, YASUHIKO
/ TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
/ FILE REFERENCE: 084335/0160
/ CURRENT APPLICATION NUMBER: US/10/094, 749
/ PRIOR FILING DATE: 2002-03-12
/ PRIOR APPLICATION NUMBER: 60/350, 435
/ PRIOR FILING DATE: 2002-01-24
/ PRIOR APPLICATION NUMBER: JP 2001-328381
/ NUMBER OF SEQ ID NOS: 3381
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 400
/ LENGTH: 2208
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-10-094-749-400

Alignment Scores:
Pred. No.: 3 65e-225 Length: 2208
Score: 1944.00 Matches: 386
Percent Similarity: 93.24% Conservative: 0
Best Local Similarity: 93.24% Mismatches: 4
Query Match: 91.48% Indels: 24
DB: 18 Gaps: 1

US-10-649-273-2 (1-414) x US-10-094-749-400 (1-2208)
QY 1 MetLeuIleLeuThrIleThrAlaGlyValPhePheLysProSerLysArgLysValIyr 20
Db 344 ATGCTAATCTTGACTAAGATGAGCGAGATTTTAAAAACATCAAAAAGAAAGTTAT 403
QY 21 GluPheLeuArgSerPheAsnPheHisPheProGlyThrLeuPheLeuHisIleValLeu 40
Db 404 GAATTTTAAAGAAAGTTTAAATTTTCATCTGGAACACTAATTTCTCATTAATAATGATTTG 463
QY 41 GlyIleGlnPheSerCysAspAspThrAlaAlaAlaValValAspGluThrGlyAsnVal 60
Db 464 GGAATTTAAACTAATTTGATGATACAGCACTGCTGTGTGATGAACCTGGAATG 523
QY 61 LeuGlyGlnAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyIleValPro 80
Db 524 TTGGGAGAAGCAATACATTTCCCAACGAAGTTCATTAATAAACAAGTGGATGTTCT 583
QY 81 ProAlaIleGlnGlnLeuHisArgGlnAsnIleGlnArgIleValGlnGlnAlaLeuSer 100
Db 584 CCAGCAGCTCAACAGCTTCAACAGAAATTTCAACAAATAGTCAAGAAAGCTCTTCT 643
QY 101 AlAserGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
Db 644 GCCAGTGAAGTCTCTCCAAAGTGAACCTTCAGCAATTCGAACTACCACTAAACAGAGACTT 703
QY 121 AlaleuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
Db 704 GCTTTAAGCCTGGAGAGTGGCTTATCACTTACCTTACAGCTGTGTGAGCAAGTAAAAAG 763
QY 141 ProPheIleProIleHisIleMetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 764 CCATTCAATCCCATTCATCATATGAGGCTCATGACATTACTAATTAAGTTGACCAATAAA 823

```

161 ValGluPheProPheLeuValLeuLeuIleSerGlyYHisCysLeuLeuAlaLeuVal 180  
Db 824 GAGAAATTTCTTTTATGTTCTTTGATTTCTGGAGGTCACGTCTGTTGGCAATTAGT 883  
Qy 181 GInGlyValSerAspPheLeuLeuGlyLeuSerLeuAspIleAlaProGlyAspMet 200  
Db 884 CAAGAGTTTCCAGATTTTCTGCTTCTTGGAAAGTCTTTGGACATAGCACAGGTGACATG 943  
Qy 201 LeuAspLeuValAlaArgArgLeuSerLeuIleLeuHisProGlyCysSerThrMetSer 220  
Db 944 CTTCAGAGGTGGCAAGAGACTTCTTAAATTAACATCCAGAGTCTCCACCATAGT 1003  
Qy 221 GlyGlyValAlaIleGluHisLeuAlaLeuGlnGlyAsnArgPheHisPheAspIleLeu 240  
Db 1004 GGTGGAAAGCCATAGAACATTTGGCCAAACAAAGAAATGATTTGATTTGACATCAAA 1063  
Qy 241 ProProLeuHisHisAlaLeuAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
Db 1064 CCTCCCTTGATCATAGCTTAATAATTTGATTTCTTTACTGAGCTTCAACACGTTACT 1123  
Qy 261 AspLeuIleIleMetLeuLeuGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 280  
Db 1124 GATTAATAATATATGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAG 1183  
Qy 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLeuArg 300  
Db 1184 TCAGCAGCAGACATTTGCTCCACAGTACAGCACACAAATGATGTCATCTTGGAAGAA 1243  
Qy 301 ThrHisArgAlaAlaLeuPheCysLeuGlnArgAspLeuLeuProGlnAsnAlaVal 320  
Db 1244 ACACATCGGGCTATTTCTGTTTGTAAAGCAGAGACTTTGATCTTCAAAAATATATGACAG 1303  
Qy 321 LeuValAlaSerGlyValAlaLeuAsnAsnPheTrpIleArgArgAlaLeuGlnIleLeu 340  
Db 1304 CTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1363  
Qy 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspGln 360  
Db 1364 ACAAAAGCAACACATGCACTTTGTTGTGCTCTCCAGACTAGTATGACATGATATGG 1423  
Qy 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
Db 1424 ATTTGATTTGCA-----TGATGCTCTTGGAGTACATATCAAAAAGAAAT 1435  
Qy 381 IleGluGlyIleArgTrpGluProLeuCysProLeuGlyValAspIleSerLeuVal 400  
Db 1436 -----TGATGCTCTTGGAGTACATATCAAAAAGAAAT 1471  
Qy 401 GlyGluAlaSerIleLeuValProGlnLeuLeuMetGluIle 414  
Db 1472 GGAGAGCTTCCATTAAGTACCAATTAATAATGAGATA 1513

RESULT 12  
US-10-723-860-7447  
; Sequence 7447, Application US/10723860  
; Publication No. US20040253606A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natsaba  
; APPLICANT: Ginsburg, Wendy M.  
; APPLICANT: Zlotnik, Albert  
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
; FILE REFERENCE: 05982, 0193, NUS01  
; CURRENT APPLICATION NUMBER: US/10/723,860  
; PRIOR FILING DATE: 2003-11-26  
; PRIOR APPLICATION NUMBER: 60/429,739  
; NUMBER OF SEQ ID NOS: 8393  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO: 7447  
; LENGTH: 2890  
; TYPE: DNA  
; ORGANISM: Homo sapiens

FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (646)..(657)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-723-860-7447  
Alignment Scores:  
Pred. No.: 5.59e-225 Length: 2890  
Score: 1944.00 Matches: 386  
Percent Similarity: 93.24% Conservative: 0  
Best Local Similarity: 93.24% Mismatches: 4  
Query Match: 91.48% Indels: 24  
Gaps: 1  
US-10-649-273-2 (1-414) x US-10-723-860-7447 (1-2890)  
Qy 1 MetLeuIleLeuThrIleThrAlaGlyValPhePheLeuProSerLeuArgLeuValIle 20  
Db 1001 ATGCTAATCTTGATTAAGTACGAGGATTTTAAACCATCAAAAGAAAGTTAT 1060  
Qy 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLeuValLeu 40  
Db 1061 GAATTTTAAGAATTTTAATTTTCACTGGAACACTAATTTCTTAATAATATGATG 1120  
Qy 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60  
Db 1121 GGAATTTGAATCTAGTTGATGATGATGACAGAGCTGCTGCTGATGATGAACTGGAAATG 1180  
Qy 61 LeuGlyGluAlaIleHisSerGlnThrGlnValHisLeuLeuThrGlyIleValPro 80  
Db 1181 TTGGAGAAAGCAATACATCTCCAAAGTGAATTAATAAAACAGGTGGATTTGCT 1240  
Qy 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
Db 1241 CCGACAGCTCAACAGCTTCAAGAGAAAATTAACAAGAAATAGTACAAAGAGCTTTCT 1300  
Qy 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLeu 120  
Db 1301 GCAGTGGAGTCTCTCAATGACCTTCAAGCAATTCACATCAATTAATAACAGAGCTT 1360  
Qy 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyLeuLeuLeu 140  
Db 1361 GCTTTAAGCTGGAGAGGCTTATCATTTAAGCTTACAGCTGGTGAAGACATTAAAAAG 1420  
Qy 141 ProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeu 160  
Db 1421 CCAATTAATCCCATTAATCAATATGAGAGCTCAAGCACTTATAGTTGACCAATAAA 1480  
Qy 161 ValGluPheProPheLeuValLeuLeuIleSerGlyYHisCysLeuLeuAlaLeuVal 180  
Db 1481 GTAGAAATTTCTTTTATGTTCTTTGATTTCTGAGGTCACTGCTGTTGGCATTTAGTT 1540  
Qy 181 GInGlyValSerAspPheLeuLeuGlyLeuSerLeuAspIleAlaProGlyAspMet 200  
Db 1541 CAAGAGTTTCCAGATTTTCTGCTTCTTGAAGAATCTTTGGACATAGCACAGGTGACATG 1600  
Qy 201 LeuAspLeuValAlaArgArgLeuSerLeuIleLeuHisProGlyCysSerThrMetSer 220  
Db 1601 CTTCAGAGGTGGCAAGAGACTTCTTAAATTAACATCCAGAGTCTCCACCATAGT 1660  
Qy 221 GlyGlyValAlaIleGluHisLeuAlaLeuGlnGlyAsnArgPheHisPheAspIleLeu 240  
Db 1661 GGTGGAAAGCCATAGAACATTTGGCCAAACAAAGAAATGATTTGATTTGACATCAAA 1720  
Qy 241 ProProLeuHisHisAlaLeuAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
Db 1721 CCTCCCTTGATCATAGCTTAATAATTTGATTTCTTTACTGAGCTTCAACACGTTACT 1780  
Qy 261 AspLeuIleIleMetLeuLeuGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 280  
Db 1781 GATTAATAATATATGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAG 1840  
Qy 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLeuArg 300

Db 1841 TCAGACACAGACATTGCTGCCACACTACAGCACAAATGCAATGCTCATCTTGAAAACA 1900  
Qy 301 ThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuProGlnAsnAsnAlaVal 320  
Db 1901 ACACATCGGCGCTATTCTCTTTTGTAGACAGAGACTTGTTACCTCAAAATATTCAGATA 1960  
Qy 321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeu 340  
Db 1961 CTGGTTGCACTCGGAGTGTCGCAAGTAACCTCTATCCGACAGACTCTGGAATTTTGA 2020  
Qy 341 ThrAsnAlaThrGlnCysThrIleuLeuCysProProArgLeuCysThrAspAsnGly 360  
Db 2021 ACAACCGACACACAGTGCATTTGTGTCTCTCCACAGACTAGCATATTAATGAGC 2080  
Qy 361 IleMetIleAlaTrpAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
Db 2081 ATATGATTGCA----- 2092  
Qy 381 IleGlnGlyIleArgTyrGlnProIysCysProLeuGlyValAspIleSerIysGlyVal 400  
Db 2093 -----TGATGTCTCTTGAGTAGACATATCAAAAGAGATT 2128  
Qy 401 GlyGluAlaSerIleIysValProGlnLeuIysMetGlnIle 414  
Db 2129 GGAGAAAGCTTCATAAAGTACCAATTTAAATGAGATA 2170

## RESULT 13

US-10-120-988-177  
; Sequence 177, Application US/10120988  
; Publication No. US20030219745A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Ren, Feiyun  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 802CON  
; CURRENT APPLICATION NUMBER: US/10/120, 988  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: 09/774, 528  
; PRIOR FILING DATE: 2001-01-30  
; NUMBER OF SEQ ID NOS: 441  
; SOFTWARE: pc\_fl\_genes Version 2.0  
; SEQ ID NO 177  
; LENGTH: 1416  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (205) .. (1305)  
US-10-120-988-177

## Alignment Scores:

Pred. No.: 1,74e-201 Length: 1416  
Score: 1747.00 Matches: 340  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.71% Mismatches: 0  
Query Match: 82.21% Indels: 0  
DB: 18 Gaps: 0

US-10-649-273-2 (1-414) x US-10-120-988-177 (1-1416)

Qy 74 LysThrGlyGlyIleValIleProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArg 93  
Db 280 AGAACGATGGGATGTTCTCCAGACGCTCAACAGCTTCAACAGAAATATTCACAGCA 339  
Qy 94 IleValGlnGlnIleLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113  
Db 340 ATAGTCAACAAAGAGCTTTTCTGCGAGTGAAGTCTCTCAAGTGAAGCTCTCAGCAATTCGA 399

Qy 114 ThrThrIleLeuProGlyLeuAlaLeuSerIleGlyValGlyLeuSerPheSerIleGln 133  
Db 400 ACTACCAATAAACAGAGACTTCTTAAAGCTGGAGGTGGGCTTATCATTTAGCTTACAG 459  
Qy 134 LeuValGlyGlnLeuIlyAspProPheIleProIleHisMetGlnAlaHisAlaLeu 153  
Db 460 CTGGTAGGACAGTTAAAGCAATTCATTCATCCATCATATGAGAGCTCATGACATT 519  
Qy 154 ThrIleArgLeuThrAsnIysValGlnPheProPheLeuValLeuLeuIleSerGlyGly 173  
Db 520 ACTATTAGGTGACCAATTAAGTAGAATTTCTTTTATGTTTGTATTTGATTTCTGAGGT 579  
Qy 174 HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyIysSerLeu 193  
Db 580 CACTGCTGTGTGCACTTATGTTCAAGAGTTTCAGATTTTCTGCTTCTGGAAGTCTTGG 639  
Qy 194 AspIleAlaProGlyAspMetLeuAspIysValAlaArgArgLeuSerLeuIleYHis 213  
Db 640 GACATAGCACACAGGTGACATGCTTGACAAAGTGCAAGAAAGACTTCTTTAAATAAACAT 699  
Qy 214 ProGlnCysSerThrMetSerGlyGlyValAlaIleGlnHisLeuAlaIysGlnGlyAsn 233  
Db 700 CCAGAGTGTCCACCATGAGTGTGGAAAGCCATGAAACATTTGGCCAAAGAAAT 759  
Qy 234 ArgPheHisPheAspIleIysProProLeuHisHisAlaIysAsnCysAspPheSerPhe 253  
Db 760 AGATTTCATTTTGAACATCAAACTCTCTTCATCATCTTAAATTTGTATTTTCTTT 819  
Qy 254 ThrGlyLeuGlnHisValThrAspIysIleIleMetIysIysGlnGlyGlyIle 273  
Db 820 ACTGCACTTCAACAGCTTACTGATTAATATATAGAAAAGAAAGAAAGGATATT 879  
Qy 274 GlnIysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMet 293  
Db 880 GAGAAAGGCAATCTGCTTCAAGACAGACATTCCTCCACATGACGACACACATG 939  
Qy 294 AlaCysHisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeu 313  
Db 940 GCATGTCATCTTGTGAAGAAACACATCGGGCTATTTCTGTTTGTAAAGACAGACTTG 999  
Qy 314 LeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaAspAsnPheTyrIle 333  
Db 1000 TTACTCTAAATATATGACATGCTGTTGATCTGTGTGTGCGCAAGTAATCTTATATC 1059  
Qy 334 ArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPro 353  
Db 1060 CGGAGAGCTCTGGAAATTTTAAACAAAGCAACAGTGCACACTTGTGTCTCTCC 1119  
Qy 354 ArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGlnArgLeuArgAla 373  
Db 1120 AGACTATGACCTGATTAATGGATTAATGATGATGATGATGATGATGATGATGATGAT 1179  
Qy 374 GlyLeuGlyIleLeuHisAspIleGlyGlyIleArgTyrGlnProIysCysProLeuGly 393  
Db 1180 GGCTTGGCATTTTACATGACATGAGAGGATCCGCTATGAAACAAATATGCTCTTGA 1239  
Qy 394 ValAspIleSerIysGlyValGlyGlnAlaSerIleIysValProGlnLeuIysMetGln 413  
Db 1240 GTAGACATATCAAAAGAAAGTTGGAAGAGCTTCATTAAGTAACCAATTAAGAAATGAG 1299  
Qy 414 Ile 414  
Db 1300 ATA 1302

## RESULT 14

US-10-067-443-23  
; Sequence 23, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED  
; TITLE OF INVENTION: SPINDAL CORD, ME-1

```
FILE REFERENCE: D0073 NP
CURRENT APPLICATION NUMBER: US/10/067,443
CURRENT FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.0
SEQ ID NO 23
LENGTH: 1526
TYPE: DNA
ORGANISM: homo sapiens
US-10-067-443-23

Alignment Scores:
Pred. No.: 2,29e-157 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 65.18% Indels: 0
Gaps: 0

US-10-649-273-2 (1-414) x US-10-067-443-23 (1-1526)

QY 148 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIleValGluPheProPheLeuVal 167
Db 1 ATGAGAGCTCATGCACTACTATTAGTTGACCAATAAGTGAATTCCTTTTATGTT 60

QY 168 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187
Db 61 CTTTGTGATTTCTGGAGGTCAGCTGCTGTGGCATTAGTTCAAGAGTTTCATATTTCTG 120

QY 188 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 207
Db 121 CTTCTTGGAAAGCTTTTGGACATGACACAGGTGACATGCTTGACAAAGTGCGAAGAAGA 180

QY 208 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 227
Db 181 CTTTCTTTAATAAACAATCCAGAGTCTCCACATGATGGTGGGAAAGCCATAGAACT 240

QY 228 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 247
Db 241 TTGGCCCAACAGAGAAATAGATTTCATTTGATCAATCAACCTCCCTGCATCATCTAA 300

QY 248 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 267
Db 301 AATTGTGATTTTCTTTTACTGACCTTCAACAGCTTACTGATAAATATATGAAAAAG 360

QY 268 GluLysGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIle 287
Db 361 GAAAAAAGAGAGATTTGAGAAAGGGGCAATCTGTCTTCAGCGACGACATTCCTGCC 420

QY 288 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 307
Db 421 ACAGTACAGACACAATGCGATGTCATCTTGTGAAAAAACAACATCGGGCTATTCGTTT 480

QY 308 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 327
Db 481 TGTAAAGAGAGAGCTGTGTAACCTCAAAATATGACATGCTGTGCTGTGCTGTC 540

QY 328 AlaSerAsnPheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 347
Db 541 GCAGATTAATCTTATATCCGAGAGCTGTGAAATTTTAAACAAGCAACAGAGGCACT 600

QY 348 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 367
Db 601 TTGTGTGCTCTCTCTCCAGACTATGCACTGATATAGCATTTATGATTCATGGAATGCT 660

QY 368 TleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgGlyGlu 387
Db 661 ATGAAAGACTACAGCTGCTGCTGGCATTTATCATGACATAGAAGCATCGGCTATGAA 720
```

```
QY 388 ProLysCysProLeuGlyValAlaAspIleSerLysGlnValGlnAlaSerIleLysVal 407
Db 721 CCAAAATGTCCTTGGAGTAGACATATCAAAAGAAAGTGGAGAGCTTCATATAAGTA 780

QY 408 ProGlnLeuLysMetGluIle 414
Db 781 CCACAATTAAATAATGGAGATA 801

RESULT 15
US-10-649-273-23
Sequence 23. Application US/10649273
Publication No. US20040043407A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
FILE REFERENCE: D0073 CNT
CURRENT APPLICATION NUMBER: US/10/649,273
CURRENT FILING DATE: 2003-08-27
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 10/067,443
PRIOR FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.2
SEQ ID NO 23
LENGTH: 1526
TYPE: DNA
ORGANISM: homo sapiens
US-10-649-273-23

Alignment Scores:
Pred. No.: 2,29e-157 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 65.18% Indels: 0
Gaps: 0

US-10-649-273-2 (1-414) x US-10-649-273-23 (1-1526)

QY 148 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIleValGluPheProPheLeuVal 167
Db 1 ATGAGAGCTCATGCACTACTATTAGTTGACCAATAAGTGAATTCCTTTTATGTT 60

QY 168 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187
Db 61 CTTTGTGATTTCTGGAGGTCAGCTGCTGTGGCATTAGTTCAAGAGTTTCATATTTCTG 120

QY 188 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 207
Db 121 CTTTGTGAAAGCTTTTGGACATGACACAGGTGACATGCTTGACAAAGTGCGAAGAAGA 180

QY 208 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValLysAlaIleGluHis 227
Db 181 CTTTCTTTAATAAACAATCCAGAGTCTCCACATGATGGTGGGAAAGCCATAGAACT 240

QY 208 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 247
Db 241 TTGGCCCAACAGAGAAATAGATTTCATTTGATCAATCAACCTCCCTGCATCATCTAA 300

QY 248 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 267
Db 301 AATTGTGATTTTCTTTTACTGACCTTCAACAGCTTACTGATAAATATATGAAAAAG 360

QY 268 GluLysGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIle 287
Db 361 GAAAAAAGAGAGATTTGAGAAAGGGGCAATCTGTCTTCAGCGACGACATTCCTGCC 420

QY 288 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 307
Db 421 ACAGTACAGACACAATGCGATGTCATCTTGTGAAAAAACAACATCGGGCTATTCGTTT 480
```

```

Qy 308 CysIysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 327
Db 481 TGTAAAGCAGAGAGACTTGTTCCTCAAAATAATGACAGTACTGGTTCATCTGGTGGTCTC 540
Qy 328 AlaSerAsnPhenylrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 347
Db 541 GCAAGTAACTTCTATATCCGACAGACTCTGGAAATTTTAAACAACGCAACACAGTCACT 600
Qy 348 LeuLeuCysProProArgLeuLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 367
Db 601 TTGTTGTCTCTCTCTCCAGACTATGCACTGATATGCAATTATGATGCAATGAGT 660
Qy 368 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGln 387
Db 661 ATTGAAGACTACGCTGCTGGCTTGGCAATTTTACATGACATAGAGGCATCCGCTATGAA 720
Qy 388 ProIysCysProLeuGlyValAspIleSerIysGlnValGlyGlnAlaSerIleLysVal 407
Db 721 CCAAAATGTCTCTTGGAGTAGACATATCAAAAGAGTTGGAGAGCTTCCATTAAGAATA 780
Qy 408 ProGlnLeuLysMetGlnIle 414
Db 781 CCACATTTAAATGAGATTA 801

```

Search completed: November 11, 2005, 03:16:23  
 Job time : 1401.75 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

# OM protein - nucleic search, using frame\_plus.p2n model

Run on: November 10, 2005, 16:52:13 ; Search time 8167.05 Seconds  
(without alignments)  
2456.267 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125 1 MLITKTAGVFKPSKRRVY.....DISKEVGASIKVPLKMEI 414

Sequence: 1 MLITKTAGVFKPSKRRVY.....DISKEVGASIKVPLKMEI 414

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 470823 segs, 2427607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n model -DEV=xlp  
-Q=/cgr21/USFPO.spool/p/US10649273/runat\_02112005\_091338\_15550/app\_query.fasta\_1.1429  
-DB=GenBdb1 -OPMT=faetap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pcp -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pic -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US10649273 @CGN 1 1 7742 @runat\_02112005\_091338\_15550 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEOUTRY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : GenBdb1:  
1: gb\_ba:\*  
2: gb\_ncg:\*  
3: gb\_in:\*  
4: gb\_on:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_pn:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_ats:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	1908	9	BC011904 Homo sapi
2	2125	100.0	1907	6	AR428803 Sequence
3	2090.5	98.4	1387	6	AR428808 Sequence
4	2090.5	98.4	1387	9	HSA295148 Homo sapi

5	2088	98.3	1245	6	AX664697 Sequence
6	2088	98.3	1820	6	AX664695 Sequence
7	1944	91.5	2208	6	AX713716 Sequence
8	1944	91.5	2208	9	AK055441 Homo sapi
9	1835	86.4	1844	10	BC058172 Mus muscu
10	1747	82.2	1416	6	AR541929 Sequence
11	1725	81.2	1546	10	BC078974 Rattus no
12	1385	65.2	1526	6	AR428809 Sequence
13	1362	64.1	1522	6	AX943991 Gallus ga
14	1273	59.9	1017	10	BC038910 Mus muscu
15	1208	56.8	1558	9	AX930963 Gallus ga
16	1193.5	56.2	8415	9	AC013468 Homo sapi
17	1186.5	55.8	14364	6	AR428807 Sequence
18	995.5	46.8	249601	2	AC114153 Rattus no
19	995.5	46.8	308652	2	AC121478 Rattus no
20	970	45.6	256751	10	AC122925 Mus muscu
21	950.5	44.7	1109	10	BC051211 Mus muscu
22	938	44.1	860	5	BX930694 Gallus ga
23	827	38.9	875	6	CQ721898 Sequence
24	751.5	35.4	121251	5	AL591593 Zebrafish
25	714.5	33.6	1576	3	AY051882 Drosophila
26	714.5	33.6	1601	6	CQ606432 Sequence
27	710	33.4	1474	3	AK113378 Ciona int
28	709	33.4	11732	5	AL672217 Zebrafish
29	700.5	33.0	1385	6	BD157102 Primer fo
30	700.5	33.0	1385	6	AX878239 Sequence
31	700.5	33.0	1385	9	AK027836 Homo sapi
32	687.5	32.4	3656	6	CQ606431 Sequence
33	687.5	32.4	14679	2	AC018262 Drosophila
34	687.5	32.4	180263	3	AC010671 Drosophila
35	687.5	32.4	207432	3	AE003513 Drosophila
36	668.5	31.5	1443	8	AY024338 Arabidops
37	668.5	31.5	1474	8	AY117283 Arabidops
38	668.5	31.5	1567	8	AY063864 Arabidops
39	662.5	31.2	1557	8	AY084577 Arabidops
40	656	29.6	1672	8	AK070912 Oryza sat
41	629	29.6	571	6	BD154975 Primer fo
42	629	29.6	571	6	AX874913 Sequence
43	618	29.1	110000	2	AP006501_08 Continuation (9) of
44	577.5	27.2	260271	1	AE017258 Wolbachia
45	568.5	26.8	333800	1	SM591792 Stenothrips

## ALIGNMENTS

RESULT 1	BC011904	1908 bp	mrna	linear	PRI 23-DEC-2003
LOCUS	BC011904				
DEFINITION	Homo sapiens O-sialoglycoprotein endopeptidase-like 1, mRNA (cDNA clone MGC:20293 IMAGE:4121450), complete cds.				
ACCESSION	BC011904				
VERSION	BC011904.2	GI:40225818			
KEYWORDS	MGC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 1908) Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, P.S., Wagner, C.M., Shenmen, C.F., Bhat, N.K., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ushed, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hult, S.W., Viallon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, B., Ketterman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,				



QY 321 LeuValAlaSerGlyValAlaSerAspPheTyrIleArgArgAlaLeuGluIleLeu 340  
Db 1064 CTGGTGGCACTCGGAGTGTGCGAAGTACTTCTATATCCGAGACCTCGGAAATTTTA 1123  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360  
Db 1124 ACAAGCGACACAGCGACTTGTGTGTCTCTCCCGAGACTAGCACTGATTAATGGC 1183  
QY 361 IleMetIleAlaTTPAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
Db 1184 ATTATGATTCAGCAATGATGATGTAAGAACTACGCTGCGCTTGGGCAATTTACATGAC 1243  
QY 381 IleGluGlyIleArgTyrGluProLeuCysProLeuGlyValAspIleSerLeuGluVal 400  
Db 1244 ATGAGAGGCATCCGCTATACCAAAATGTCTTGGAGTAGACATATCAAAAGAAATT 1303  
QY 401 GlyGluAlaSerIleLeuValProGlnLeuLeuMetGluIle 414  
Db 1304 GGAGAAAGCTTCATTAAGTACCAATTAATAATGAGATA 1345  
RESULT 2  
AR428803 2197 bp DNA linear PAT 18-DEC-2003  
LOCUS AR428803 Sequence 1 from patent US 6642041.  
ACCESSION AR428803  
VERSION AR428803.1 GI:40188589  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2197)  
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.  
TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 6642041-A 1 04-NOV-2003;  
FEATURES  
LOCATION/Qualifiers  
1..2197  
/organism="Unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Alignment Scores:  
Pred. No.: 7,37e-171 Length: 2197  
Score: 2125.00 Matches: 414  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
DB: 6  
US-10-649-273-2 (1-414) x AR428803 (1-2197)  
QY 1 MetLeuIleLeuThrIleThrAlaGlyValAlaPhePheLeuProSerLeuArgLeuValTyr 20  
Db 231 ATGCTAATCTTACCTAAGACTGACGAGAGTTTCTTTTAAACATCAAAAAGAAAGTTTAT 290  
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40  
Db 291 GAATTTTAAAGAGTTTAAATTTTCACTCTGGAACACTATTTCTTCAATAAATAGATTTG 350  
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaAlaValAlaAspGluThrGlyAsnVal 60  
Db 351 GGAATTTGAAACTAGTGTGATGATACAGCAGCTGCTGTGGATGAACCTGGAATGTGC 410  
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLeuSerThrGlyIleValPro 80  
Db 411 TTGGAGAGAGCAATTCATTTCCAAACTGAAGTTCAATTTAAAAACAGGTGGATTTGCTCT 470  
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
Db 471 CCAGAGAGCTCAACAGCTTCAACAGAAATATTTCAACGAATGATACAAAGAGCTCTTTCT 530  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLeu 120  
Db 531 GCCAGTGGAGTCTCTCCAAAGTACCTCTCAGCAATTGCAACTACATAAACAAGAGACTT 590

QY 121 AlaLeuSerLeuGlyValAlaGlyLeuSerPheSerIleuGlnLeuValGlyGlnLeuLeuGly 140  
Db 591 GCTTTAAGCCTGGAGAGGGCTTATCATTTAGCTTACAGCTGGTAGAGCAAGTTAAAAAG 650  
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeu 160  
Db 651 CCAATCATTCCTCAATTCATATGAGAGCTCAATGCACTTCACTATTTAGGTGACCAATTA 710  
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
Db 711 GTAGAAATTCCTTTTATAGTCTTTTATGATTTCTGAGAGTCACTGCTGTGGCAATTTGTT 770  
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyValSerLeuAspIleAlaProGlyAspMet 200  
Db 771 CAAGAGATTTCAGATTTTCTGCTTTGGAAGCTTTTGACATACGACCAAGTGCATGATG 830  
QY 201 LeuAspLeuValAlaArgArgLeuSerLeuIleHisIleProGluCysSerThrMetSer 220  
Db 831 CTGACAAAGGTGGCAAGAACCTTTCTTATATAAATCATCCAGTGTCTCCACCATAGAT 890  
QY 221 GlyGlyValAlaIleGluHisLeuAlaLeuAlaGlyValAsnArgPheHisPheAspIleLeu 240  
Db 891 GGTGGGAAAGCCATAGAACATTTGGCCAAACAAGAAATAGATTTCAATTTTGAATCAA 950  
QY 241 ProProLeuHisIleAlaLeuAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
Db 951 CTCCTCTTGATCATGCTAATAAATTTGATTTTCTTTACTGACCTTCAACAGCTTACT 1010  
QY 261 AspLeuIleIleIleMetLeuValGluValGluGluGlyIleGluValGlyGlnIleLeuSer 280  
Db 1011 GATTAATAATATAATGAAAAAGAAAAAGAGAAAGTTTGAAGAGGGCAAAATCTGTCT 1070  
QY 281 SerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLeuArg 300  
Db 1071 TCAGCAGACAGACTTGTGCCACAGTACAGACACAAATGGCATGTCACTTGTGAAAGA 1130  
QY 301 ThrHisArgAlaIleLeuPheCysGlyGlnArgAspLeuProGlnAsnAlaVal 320  
Db 1131 ACACATCGGCTATTTCTGTTTGTAAAGCAGAGACTGTATCTCAAAATAATAGCAGTA 1190  
QY 321 LeuValAlaSerGlyValAlaSerAspPheTyrIleArgArgAlaLeuGluIleLeu 340  
Db 1191 CTGGTTGCATCTGTGTGTGTCGCAAGTACTTCTATATCCGAGACCTCGGAAATTTTA 1250  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360  
Db 1251 ACAAGCGACACAGCTGACTTGTGTGTCTCTCTCCAGACTATGCACTGATTAATGGC 1310  
QY 361 IleMetIleAlaTTPAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
Db 1311 ATTATGATTCAGTGAATGATATTTGAAGACTAGCTGCTGGCAATTTTACATGAC 1370  
QY 381 IleGluGlyIleArgTyrGluProLeuCysProLeuGlyValAspIleSerLeuGluVal 400  
Db 1371 ATGAGAGGCATCCGCTATGAAACCAAAATGTCTCTTGAAGTAGACATATCAAAAGAGTT 1430  
QY 401 GlyGluAlaSerIleLeuValProGlnLeuLeuMetGluIle 414  
Db 1431 GGAGAAAGCTTCATTAAGTACCAATTAATAATGAGATA 1472  
RESULT 3  
AR428808 1387 bp DNA linear PAT 18-DEC-2003  
LOCUS AR428808 Sequence 21 from patent US 6642041.  
ACCESSION AR428808  
VERSION AR428808.1 GI:40188594  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1387)  
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.

TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 662041-A 21 04-NOV-2003;  
FEATURES location/Qualifiers  
SOURCE 1..1387  
/organism="unknown"  
/mol\_type="genomic DNA"

## ORIGIN

## Alignment Scores:

Score:	3.54e-168	Length:	1387
Percent Similarity:	2090.50	Matches:	412
Best local Similarity:	93.85%	Conservative:	0
Query Match:	98.38%	Mismatches:	2
		Indels:	25
		Gaps:	1

US-10-649-273-2 (1-414) x AR428808 (1-1387)

```

Oy      1 MetLeuIleLeuThyThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
Db      24 ATGCTAACTTACATACAGCTGAGAGCTTTTAAACATCAAGAAAGAAAGTTAT 83
Oy      21 GluPheLysArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysValLeu 40
Db      84 GAATTTTAAAGAGTTTATTTTCTGAAACACTATTCTCATATAAATAGTATTG 143
Oy      41 GYIIEGLIuThSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60
Db      144 GGAATTTGAAACTGATGTGATATACAGACGCTGCTGTGATGAAACTGGAAATGTG 203
Oy      61 LeuGlyValAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyGlyIleValPro 80
Db      204 TTGGAGAAAGCAATATACATTCCTCAAACTGAAGTTCAATTAACAGAGTGGAGTTGCTCT 263
Oy      81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Db      264 CCAGCAGCTCAACAGCTTCAACAGAGAAATATTCAAGATATATACAGAAAGCTCTTCT 323
Oy      101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
Db      324 GCCAGTGGAGTCTCTCCAGTGCCTCTCAGCAATTGCAACTACATMAAACAGAGACTT 383
Oy      121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140
Db      384 GCTTTAAGCTTGGAGATGGCTTATCATTTAGCTTACAGCTGGTAGCAGTTAAATAAG 443
Oy      141 ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db      444 CCATTCATTCCTCATTCATATAGAGGCTCATGCACTTATTTAGGTTGACCAATMAA 503
Oy      161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db      504 GTAGAAATTTCTTTTATGTTCTTTGATTTCTGAGAGTCACTGCTGTGGCATTTAGTT 563
Oy      181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
Db      564 CAAGGAGTTTCAGATTTTCTGCTTCTTGAAGCTTTTGAACATAGCACAGGTCATG 623
Oy      201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
Db      624 CTTGACAAAGTGGCAAGAGACTTTCTTTAATTAATAACATCCAGAGTCTCCACCATAGT 683
Oy      221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
Db      684 GGTGGGAAAGCCATAGAGCATTTGGCCAAACAGAAATAGATTTCATTTTGAACATCAA 743
Oy      241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db      744 CCTCCCTTGATCATGCTAATAAATTTGATTTTCTTTTACTGACATTCAACAGTTACT 803
Oy      261 AspLysIleIleMetLysLysGlyLysGlnGlyIle----- 273
Db      804 GATTAATAATATATGAAAAAGAAAAAGAGAAAGTATATTTCTTAATTAAGTAAAGTTGAA 863

```

```

Oy      274 -----GluLys 275
Db      864 CAGATAAATATATCTGGATTTGCTCTAAATAAGCTGCTCATTTCTCAGGTATAGAGAG 923
Oy      276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 295
Db      924 GGCCAAATCTGCTTCTCAGACAGACATTCCTGCCACAGTACGACACATAGGACATG 983
Oy      296 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnAspLeuLeuPro 315
Db      984 CATCTGTGAAAGAACACATCCGGCTATTCTGTTTGTAAAGACAGAGACTTTTACT 1043
Oy      316 GlnAsnAsnAlaValLeuValAlaAspGlyValAlaAspAspPheTyrIleArgArg 335
Db      1044 CAAATATATCCAGTACGTTGATCTGGTGCTGCTCCAGTACACTTATATCCGACGA 1103
Oy      336 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 355
Db      1104 GCTCTGAAATTTTAAACAAACGCAACAGTGCCTTGTGTCTCTCCCTCCAGACTA 1163
Oy      356 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 375
Db      1164 TGACCTGATATATGCTATTATGATTCATGAGATGTAATTGAAGACTACAGTGGCTTG 1223
Oy      376 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 395
Db      1224 GGCATTTTACATGACATAGAAAGCATCCGCTATGACCAAAATGCTCTTGAGATAGAC 1283
Oy      396 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
Db      1284 ATATCAAAAGAGTTTGAGAAAGCTTCATATAAAGTACACAAATTAAATGAGATA 1340

RESULT 4
HSA295148      1387 bp      mRNA      linear      PRI 30-OCT-2000
LOCUS      Homo sapiens mRNA for putative sialoglycoprotease type 2.
DEFINITION      AJ295148
ACCESSION      AJ295148.1 GI:11071726
VERSION      metallopeptidase; sialoglycoprotease.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE      1
AUTHORS      Chen,J.M., Fortunato,M. and Barrett,A.J.
TITLE      Cloning and sequencing of a second human putative
JOURNAL      sialoglycoprotease homologue
REFERENCE      2 (bases 1 to 1387)
AUTHORS      Chen,J.M.
TITLE      Direct Submission
JOURNAL      Submitted (27-OCT-2000) Chen J.M., MRC Molecular Enzymology
Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT,
UNITED KINGDOM

FEATURES
source      location/Qualifiers
1..1387
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/feature_type="utermus"
24..1343
/function="metallopeptidase of family M22"
/codon_start=1
/product="putative sialoglycoprotease type 2"
/protein_id="CAC14666.1"
/db_xref="GI:11071727"
/db_xref="GOA:Q9H4B0"
/db_xref="UniProt/TREMBL:Q9H4B0"
/db_xref="SCDPTAAVNDENGLGEAHSQTEVHLKTKGIVPPAQQOIAHRENIQRIQVEALSAS
GVSPSDSAIAITIKPGALSLGVGSLQVGLKXKPFIPIHMEAHALITRLTNK
VEPFPVILLISGHCILLALVGVSDPFLLLGKSLDIAPDMDKVARRLSLIKHPCST

```

MSGGAIEHLAKOGRPHFDIKPEPLHAKNCDFSTGLOHTVDKIMKKEKEGFI  
SKVDIINPGLCKIAHFCREKQIILSSADIAITVHTACHVETHTAIIFFCK  
QRDLDPONNAIVASGVASNFYIRAHILTNATQCTILCPPLCTDNGIMIAWNG  
IERLRGIGIILHDEGIRYEPKCPGVDSKEVGASIKVPOLKMEI"

## ORIGIN

## Alignment Scores:

Pred. No.:	3,54e-168	Length:	1387
Score:	2090.50	Matches:	412
Percent Similarity:	93.85%	Conservative:	0
Best Local Similarity:	93.85%	Mismatches:	2
Query Match:	98.38%	Indels:	25
DB:	9	Gaps:	1

US-10-649-273-2 (1-414) x HSA295148 (1-1387)

QY 1 MetLeuIleLeuThrThyThraIaGlyValPhePheLeuProSerLySArgLySValTyr 20  
DB 24 ATGCTAACTTGACCTAAGACTGAGAGATTTTAAACATCAAAAAGAAAGTTTAT 83  
QY 21 GluPheLeuArgSerPheAsnBheIspProGlyThrLeuPheLeuHisLySIIeValLeu 40  
DB 84 GAATTTTAAAGAGTTTAAATTTTCATCTGAAACCTATTTCTTCAATAAATAGTATTG 143  
QY 41 GlyIIEgluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60  
DB 144 GGAATTTGAACCTAGTGTGATGATACAGCACTGCTGTGTGATGAACCTGAAATGTG 203  
QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisLeuLeuThrGlyGlyIIEValPro 80  
DB 204 TTGGAGAAAGCAATACATTTCCAAACCTGAAGTTCATTAAACAGGTGGAGTTGTTCT 263  
QY 81 ProAlaIaGlnGlnLeuHisArgGluAsnIIeGlnArgIIeValGlnGluAlaLeuSer 100  
DB 264 CCACAGACTCAACAGCTTCACAGAAATAATTCACAGAAATGACAAAGACTCTTTCT 323  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
DB 324 GCCAGTGAGTCTCTCCAAAGTGAACCTCTCAGCAATGCAATCAATAAACAGAGACTT 383  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLyS 140  
DB 384 GCTTTAACCTGGAGGTGGCTTATCATTTAGCTTACCTGAGTGGAGAGTTAAAAAG 443  
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLy 160  
DB 444 CCATTTCATTTCCCATTCATCATATGAGGCTCATGCACTTACTATTTAGTTGACCAATTA 503  
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
DB 504 GTAGAAATTTCTTTTATTGTTCTTTTGAATTCCTGAGGTCACGTGCTGTTGCAATTAGTT 563  
QY 181 GlnGlyValSerAspPheLeuLeuGlyLySLeuAspIIeAlaProGlyAspMet 200  
DB 564 CAAGAGATTTCAATTTCTGCTCTTGAAAGCTTTTGACATATGCAACAGTGCACATG 623  
QY 201 LeuAspLySValAlaArgArgLeuSerLeuIleLySHisAspGlnCysSerThrMetSer 220  
DB 624 CTTTACAAAGTGGCAAGAACTTTCTTTAATAAACAATCCAGAGGCTCCCAACATGAGT 683  
QY 221 GlyGlyLySAlaIleGluHisLeuAlaLySIIeGlnGlyAsnArgPheHisAspIIeLyS 240  
DB 684 GGTGGAAAGCCATAGACATTTTGCCCAACAAAGAAATAGATTTTCAATTTTACATCAA 743  
QY 241 ProProLeuHisHisAlaLySAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB 744 CCTCCCTGCATCATGCTAAAAATGTGATTTTCTTTTACGCACTTCAACACGTTACT 803  
QY 261 AspLySIIeIleMetLySIIeGluGlyGluGlyGlyIIE----- 273  
DB 804 GATTAATAATTAATGAAGAAAGAAAGAGAGATATTTTCTAATTAGTAAGTTGAA 863  
QY 274 -----GluLyS 275

DB 864 CAGATTAATATTTCTGCGATTGCTCTTAAAAATACCTGCTCAFTTTCGAGGTATGAGAG 923  
QY 276 GlyGlnIleLeuSerSerAlaAlaAspIIeAlaAlaThrValGlnHisThrMetAlCys 295  
DB 924 GGGCAATCTGCTTTCAGCAGCAGACATTTGCTGCCACATGACGACACAAATGGCATGT 983  
QY 296 HisLeuValLySArgThrHisAlaGAlaIleLeuPheCysLySIIeGlnAspLeuPro 315  
DB 984 CATCTGTGAAAGAACACATCGGCTATTTCTGTTTGTAAAGCAGAGACTTGTACT 1043  
QY 316 GluAspAspAlaValLeuValAlaSerGlyGlyValAlaSerAsnBheThrIleArgArg 335  
DB 1044 CAAAATATATCAGTACTGTTGATGCTGTGGGTGTCGAAGTAACTTTATATATCCGAGA 1103  
QY 336 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 355  
DB 1104 GCTTGGAAATTTTAAACAAACGCAACAGCTGCACTTTGTTGTCTCTCCACAGACTA 1163  
QY 356 CysThrAspAsnGlyIIEMetIleAlaTrpAsnGlyIIEGluArgLeuArgAlaGlyLeu 375  
DB 1164 TGCACGTATATGATGATATGATGATGATGATGATGATGATGATGATGATGATGATG 1223  
QY 376 GlyIIEleuHisAspIIeGluGlyIIEArgTyrGluProLySAspProLeuGlyValAsp 395  
DB 1224 GGCATTTTACATGACATAGAGGCAATCCGCTATGAACCAAAATGCTCTTGGAGTAC 1283  
QY 396 IleSerIleGlyValGlyGlyAlaSerIleLySValProGlnLeuLySMetGluIle 414  
DB 1284 ATATCAAAAGAGTTGAGAGAGCTTCCATTAAGATCCACATTAATAATGAGATA 1340  
RESULT 5  
AX664697 1245 bp DNA linear PAT 22-MAR-2003  
LOCUS AX664697  
DEFINITION Sequence 6 from Patent WO2074960.  
ACCESSION AX664697  
VERSION AX664697.1 GI:29164457  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Leiby,K.R., Kapeller-libermann,R. and Glucksmann,M.  
TITLE 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions  
of human proteins and uses thereof  
JOURNAL Patent: WO 02074960-A, 6 26-SEP-2002;  
Millennium Pharmaceuticals, Inc. (US)  
FEATURES  
source 1. 1245  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Alignment Scores:  
Pred. No.: 5.03e-168 Length: 1245  
Score: 2088.00 Matches: 407  
Percent Similarity: 99.03% Conservative: 3  
Best Local Similarity: 98.31% Mismatches: 4  
Query Match: 98.26% Indels: 0  
DB: 6 Gaps: 0  
US-10-649-273-2 (1-414) x AX664697 (1-1245)  
QY 1 MetLeuIleLeuThrThyThraIaGlyValPhePheLeuProSerLySArgLySValTyr 20  
DB 1 ATGCTAACTTGACCTAAGACTGAGAGATTTTAAACATCAAAAAGAAAGTTTAT 60  
QY 21 GluPheLeuArgSerPheAsnBheIspProGlyThrLeuPheLeuHisLySIIeValLeu 40  
DB 61 GAATTTTAAAGAGTTTAAATTTTCATCTGAAACCTATTTCTTCAATAAATAGTATTG 120

Oy	41	GLYLLEGIUPTHRSerCyASAPSPDThAlaAlaValaValaASPgiUthngiValaVal	60
Db	121	GGAAITGAATAGTGTGATGATGAATACAGACAGCTCTGGTGATGAAATCGAANAATG	180
Oy	61	LeuGIyGuaAlaIleHisSerGlnThrGluValHisLeuYsThngiYIleValPro	80
Db	181	TTGGAGAGACAAATACATTCGCCAAATCGAATTCATTTAAAAACAGGTGGATTGTTCT	240
Oy	81	ProAlaIaGInGlnLeuHisArgGluAsnIleGlnArgIleValGInGluAlaLeuSer	100
Db	241	CCACAGCTCAACAGCTTCACAGAGAAATATTCACAGAAATGTAACAAGAGCTCTTCT	300
Oy	101	AlaSerGlyValaSerProSerAspLeuSerAlaIleAlaThrTrpIleYsProGlyLeu	120
Db	301	GCCAGTGAAGTCTCCCAAGTACCTCCAGCAATGCACTACCAATTAACCAAGAGACTT	360
Oy	121	AlaLeuSerLeuGIyValGIyLeuSerPheSerLeuGInLeuValGIyGInLeuYs	140
Db	361	GCTTTAACCTCGGAGTGGCTTATCATTAGCTTACAGCTGGTAAGACATTAATAAG	420
Oy	141	ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnYs	160
Db	421	CCATTGATTCACATTCATATGAGAGCTCATGACTTATATAGGTTGACCAATAAA	480
Oy	161	ValGluPheProPheLeuValLeuLeuIleSerGlyHisCySLeuLeuAlaLeuVal	180
Db	481	GTAGAAATTCCTTTTAAATTCCTTTAAATTCCTGAGAGTCACTGTCGTGGCATAGTT	540
Oy	181	GInGIyValaSerAspPheLeuLeuGIyYsSerLeuAspIleAlaProGIyAspMet	200
Db	541	CAAGAGTTTACAGATTTTCTGCTTCTTGAAAGCTTTTGACATACACAGGTGACATG	600
Oy	201	LeuAspYsValaIaArgArgLeuSerLeuIleYsHisProGIyCySerThrMetSer	220
Db	601	CTTGACAGGTGGCAAGAACTTCTTTAATAAAACATCCAGAGTCTCCACATGAGT	660
Oy	221	GlyGIyYsAlaIleGlnHisLeuAlaYsGInGlyAsnArgPheHisPheAspIleYs	240
Db	661	GGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATTAATTCATTTGACATCAAA	720
Oy	241	ProProLeuHisHisAlaYsAsnCyAspPheSerPheThngiYLeuGlnHisValThr	260
Db	721	CCTCCCTTGATCATGCTTAATAATTTGATTTTCTTTTACTGACCTTCAACACGTTACT	780
Oy	261	AspYsIleIleMetYsLeuGIyYsGluGIyYIleGluYsGIyGlnIleLeuSer	280
Db	781	GATAAAATTAATAGAAACAGGAAACAAAGAGAAAGSTA TTGAGAAAGGCCAATCTGCT	840
Oy	281	SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCyHisLeuValIleArg	300
Db	841	TCAGCAGACAGCATTTGCTGCCACAGTACAGCAACAATGGCATTCCTGTGTAAAGA	900
Oy	301	ThrHisArgAlaIleLeuPheCySLeuArgAspLeuLeuProGlnAsnAsnAlaVal	320
Db	901	ACACATCGGCTATTTCTGTTTGTAAAGAGAGACTTGTACCTCAAAAATTAATGACGA	960
Oy	321	LeuValAlaSerGIyGIyValaIaSerAsnPheYIleArgArgAlaLeuGlnIleLeu	340
Db	961	CTGGTTGATCTGGTGGTGTGCAAGTAACTTATATCCGACAGAGCTCTGGAAATTTTA	1020
Oy	341	ThrAsnAlaThrGlnCySThrIleLeuCySProProArgLeuCySThrAspAsnGIy	360
Db	1021	ACAAACGCAACACAGTGCATTTGTGTGTCTCTCCACAGCTATGACATGATATATGC	1080
Oy	361	IleMetIleAlaThrAsnGIyIleGluYsArgLeuArgAlaGlyLeuGIyIleLeuHisAsp	380
Db	1081	ATTATGATTTGCATGGAAATGTATTAAGACTTACGTGCTGGCTTTGAGATTTTACATAC	1140
Oy	381	IleGIyGIyIleArgTYrGIyProYsCySProLeuGIyValaAspIleSerIySGLuVal	400
Db	1141	ATAGAAAGCATCCGCTATGTAAACAAAAATGTCTCTTTGAGTACACATATCAAAAGAGTT	1200
Oy	401	GIyGluAlaSerIleYsValaProGlnLeuYsMetGluIle	414

Db	1201	GGAGAACCTTCATTAAGTACCACTTAATAATGAGATA	1242
RESULT 6	AX664695	1820 bp	DNA
LOCUS	AX664695	Sequence 4 from Patent WO2074960.	linear
DEFINITION	AX664695		
ACCESSION	AX664695.1	GI:29164455	
VERSION			
KEYWORDS			
ORGANISM	Homo sapiens (human)		
SOURCE	Homo sapiens		
REFERENCE	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euleleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	Leibyl,K.R., Kapeller-libermann,R. and Glucksmann,M.		
TITLE	38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions of human proteins and uses thereof		
JOURNAL	Patent: WO 02074960-A 4 26-SEP-2002; Millennium Pharmaceuticals, Inc. (US)		
FEATURES	Location/Qualifiers		
source	1..1820		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
	146..1390		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAD80044.1"		
	/db_xref="GI:29164456"		
	/translation="MLITKTAGVFPFKSRKRYEFLRSEFNHFGTLFLHKIVLGIETGSCDTPAAVVDENGTGVAIHSGTVEHLKMGGLVPPAOLKRENIORIVQALSLASGVSSDSIAITTKIPGLASLSTGSLFSIQVQGLKKPPIPHHMEAHALTITLTKRVEFPVLVLLISGGHCLTALVQGVDFLLGLSLDIAGDMIDKARKSLIKHECSTMSGKAIHLAKQGNRFHFDIKPELTHAKKCDSEFSLQHTVDKNNENRKOEBGIEKGQIESGADIALATVQHTMAACHLVKTRTHRAILFFCKORDLLPQNNATLVASGGVASFYIRPAELIINATGCTLLCPPLCTDNGIMIAWNGIERLRAGILHDIEGIRYEPKCPDLPVDSIKVEGASIVPQLKKEI"		
CDS			
Alignment Scores:	8.14e-168	Length: 1820	
Pred. No.:	2088.00	Matches: 407	
Score:	99.03%	Conservative: 3	
Percent Similarity:	98.31%	Mismatches: 4	
Best Local Similarity:	98.26%	Indels: 0	
Query Match:	6	Gaps: 0	
DB:			
US-10-649-273-2 (1-414) x AX664695 (1-1820)			
QY	1	MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerIleArgIleValTyr	20
Db	146	ATGCTAACTTGCATTAAGTACCACTTAATAATGAGATA	205
QY	21	GIuPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu	40
Db	206	GAATTTTAAAGAGTTTAAATTTTCATCTGGACACATAATTTCTCATTAATAATGATTG	265
QY	41	GIyIleGIuThrIleSerCysAspAspThrAlaAlaValIleAspGIuThrGIyAsnVal	60
Db	266	GGAATTGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	325
QY	61	LeuGIyGIuAlaIleHisSerGIuThrGIuValHisLeuGIyThrGIyGIyIleValPro	80
Db	326	TGGGAGAGCAATACATTTCCCAACCTGAAGTTTCAATTTTAAACAGGAGGAGTTTCT	385
QY	81	ProAlaIleAGInGlnLeuHisArgGIuAsnIleGlnArgIleValAGInGlnAlaLeuSer	100
Db	386	CCAGAGGCTCAACACACTTCAACAGAAATATTCACAGAAATATTCACAGAGGCTTTCT	445
QY	101	AlaSerGIyValSerProSerAspLeuSerAlaIleAlaIleThrThrIleLysProGIyLeu	120
Db	446	GCGATGAGATGCTCTCCCAAGTACCTCTCAGCAATTTGACATACATTAACACAGACTT	505



QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuValGly 140  
 Db 506 GCTTAAAGCTGGAGTGGCTTATCATTTAGCTTACAGCTGTAGGACAGTTAAAAAG 565  
 QY 141 ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAspLeu 160  
 Db 566 CCATTCACTCCATTCACTATATGAGGCTCATGACCTTACATATAGGTGACCAATAAA 625  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 Db 626 GTAGAAATTTCTTTTATTGTTCTTTGATTTCGAGAGTCACGTCTGTGGCATTAAGT 665  
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
 Db 686 CAAGAGTTTCAGATTTCCTGCTTCTTGGAAGTCTTTGGACATAGACACAGCTGACATG 745  
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220  
 Db 746 CTTGACAAAGTGGCAAGAGACTTTCTTTAATAAAAATCCAGAGTCTCCACCATGAGT 805  
 QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
 Db 806 GGTGGGAAAGCCATAGAACATTTGGCCAAACAAAGAAATGATTTCACTTTTGACATCAA 865  
 QY 241 ProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGlyLeuGlnHisValThr 260  
 Db 866 CTTCCCTTGATCATCTGCTAAATAATGTGATTTTCTTTTACTGGAATTCACACAGTTACT 925  
 QY 261 AspLysIleIleMetLysLysGlyGlyGlyGlyGlyIleGlyLysGlyGlnIleLeuSer 280  
 Db 926 GATAAAATTAATGAAAAACAGAAACAAAGAAAGTAATGAAAGGGGCAAAATCCGTCT 985  
 QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
 Db 986 TCAGACAGACATGCTGCTCCACAGTACGACACATGCACTGTCCTTGGAAGAAG 1045  
 QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAlaVal 320  
 Db 1046 ACACATCGGGCTATTCTGTTTGTGTAAGCAGAGAGACTTGTACTCAAAATTAATGCACT 1105  
 QY 321 LeuValAlaSerGlyGlyValAlaAsnAsnAspPheTyrIleArgArgAlaLeuGlnIleLeu 340  
 Db 1106 CTGTTGATCATCTGCTGCTGCTCCAGTAACTTCTATATCCGACAGACTCGAATTTTA 1165  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360  
 Db 1166 ACAACGCAACACAGTGCATTTGTGTCTCTCCCTCCACAGCTAGCACTGATTAATGCG 1225  
 QY 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 Db 1226 ATTATGATTCATGCAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 1285  
 QY 381 IleGluGlyIleArgTyrGlyLeuProLysCysProLeuGlyValAspIleSerLysGlyVal 400  
 Db 1286 ATAGAGGATCGCTGATGAAACCAAAATGTCTCTTGAGTGAATATCAAAAGAAAGT 1345  
 QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
 Db 1346 GGAAGAGCTTCATTAAGTACCAATTAATAAATGAGATA 1387  
 RESULT 7  
 AX713716 2208 bp DNA linear PAT 15-APR-2003  
 LOCUS AX713716  
 DEFINITION Sequence 400 from Patent EPI293569.  
 ACCESSION AX713716  
 VERSION AX713716.1 GI:29888642  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1

AUTHORS Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. T., Isono, Y., Hiro, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuno, Y.  
 TITLE Full-length cDNAs  
 JOURNAL Patent: EP 1293569-A 400 19-MAR-2003;  
 Helix Research Institute (JP) ; Research Association for Biotechnology (JP)  
 FEATURES  
 source location/Qualifiers  
 1..2208  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,836-155 Length: 2208  
 Score: 1944.00 Matches: 386  
 Percent Similarity: 93.24% Conservative: 0  
 Best Local Similarity: 93.24% Mismatches: 4  
 Query Match: 91.48% Indels: 24  
 DB: Gaps: 1  
 US-10-649-273-2 (1-414) x AX713716 (1-2208)  
 QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
 Db 344 ATGCTAATCTTGACTAAGATGCGAGAGTTTATTAACCATCAAAAGAAAGTTAT 403  
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLysLysIleValLeu 40  
 Db 404 GAATTTTAAAGATTAAATTTTATTCATCTGGAACATATTTCTTCAATAAATGATATG 463  
 QY 41 GlyIleGluHisSerCysAspAspThrAlaAlaValAlaProGluThrGlyAsnVal 60  
 Db 464 GGAATGAAACCTAGTGTGATGATACAGCAGCTGCTGTGATGAAGACTGAAATGTCG 523  
 QY 61 LeuGlyValAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyGlyIleValPro 80  
 Db 524 TTGGAGAACCAATACACTTCCCAACTGAAAGTTATTTAAAAACAGTGGAGTTGCTCT 583  
 QY 81 ProAlaAlaGlnIleLeuHisArgGlyAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 Db 584 CCAAGCTCAACAGCTTCCACAGAAATATTCAGAAATGATCAAAAGAGCTCTTCT 643  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
 Db 644 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATTCACATCAATAAACAGAGACT 703  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140  
 Db 704 GCTTAAAGCTGGAGAGGGCTTATCATTTAGCTTACAGCTGAGTGAAGAGTTAAAAAG 763  
 QY 141 ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAspLys 160  
 Db 764 CCATTCACTCCATTCACTATATGAGAGCTCATGCACTTACATATAGGTGACCAATAAA 823  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 Db 824 GTAGAAATTTCTTTTATTGTTCTTTGATTTCGAGAGTCACGTCTGTGGCATTAAGT 883  
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
 Db 884 CAAGAGTTTCAGATTTCCTGCTTCTTGGAAGTCTTTGACATAGCACCAAGTGAATG 943  
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220  
 Db 944 CTTGACAAAGTGGCAAGAGACTTCTTTAATTAACATCAAGAGTCTCCACCATGAGT 1003  
 QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
 Db 1004 GGTGGGAAAGCCATAGAACATTTGGCCAAACAAAGAAATGATTTCACTTTTGACATCAA 1063



Oy 241 ProProLeuHisSHsAlaLysAsnCyAspPheSerPheThrGlyLeuGlnHisValThr 260  
 Db 1064 CTTCCCTTCGATCATGCTAAATATTTGATTTTCTTTTACTGAGCTTCAACAGTTACT 1123  
 Oy 261 AspLysIleIleIleLeuPheGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 280  
 Db 1124 GATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1183  
 Oy 281 SerAlaAlaAspIleAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 300  
 Db 1184 TCACACAGACGACATTCGTCACACGACGACACACGACGACATTCGTCACGACGACGAC 1243  
 Oy 301 ThrHisArgAlaIleLeuPheCysGlyGlnArgAspLeuProGlnAsnAlaVal 320  
 Db 1244 ACAATCGGCGCTATTCGTTTGTGAAGCAGACAGACTTTCCTTAATAATTAATGACAGA 1303  
 Oy 321 LeuValAlaAlaSerGlyGlyValAlaAlaSerAsnAspPheTyrIleArgArgAlaLeuGlnIleu 340  
 Db 1304 CTGGTTCATCTGGTGGTGTGTCGCAAGTAACCTCTGATTCGCGACGACGCTCGAATAATTTTA 1363  
 Oy 341 ThrAsnAlaThrGlnCysThrIleuLeuCysProProProArgLeuCysThrAspAsnGly 360  
 Db 1364 ACAACGCAACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1423  
 Oy 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 Db 1424 ATATTGATTGCA----- 1435  
 Oy 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerGlyVal 400  
 Db 1436 -----TGATGTCCCTTGAGTGTGACATCAACAAAGAGTT 1471  
 Oy 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGlnIle 414  
 Db 1472 GGAGAAAGCTTCATAAAGTACACATTAATAATGAGATA 1513

RESULT 8  
 AK055441  
 LOCUS  
 DEFINITION Homo sapiens CDNA FLJ30879 fls, clone FEBRA2004592, highly similar to Homo sapiens mRNA for putative sialoglycoproteinase type 2.  
 ACCESSION AK055441  
 VERSION AK055441.1 GI:16550166  
 KEYWORDS oligo capring; fls (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R., Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H., Sekine,M., Oabayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y., Nagahata,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M., Shiratori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T., Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H., Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M., Hirooka,S., Chiba,Y., Iehida,S., Ono,Y., Takiguchi,S., Watanabe,S., Yoshida,M., Horita,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A., Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuki,H., Oshima,A., Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T., Shiohara,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S., Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H., Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T., Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M., Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hiro,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsunura,K.,

Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T., Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J., Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K., Nagase,T., Nomura,N., Kikuchi,H., Masuo,Y., Yamashita,R., Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S. Complete sequencing and characterization of 21,243 full-length human cDNAs  
 JOURNAL  
 PUBMED  
 REFERENCE  
 AUTHORS  
 Nishi,T., Nakagawa,S., Senoh,A., Mizoguchi,H., Inagaki,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahata,K., Masuo,Y., Nagai,K. and Isogai,T. NEDO human cDNA sequencing project  
 Unpublished  
 3 (bases 1 to 2208)  
 REFERENCE  
 JOURNAL  
 TITLE  
 AUTHORS  
 Isogai,T., Otsuki,T. and Sugiyama,T.  
 Direct Submission  
 Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.  
 FEATURES  
 source  
 location/Qualifiers  
 1..2208  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="FEBRA2004592"  
 /cissue\_type="brain"  
 /clone\_lib="FEBRA2"  
 /dev\_stage="fetus"  
 /note="cloning vector: pMB18SFL3"  
 344..1438  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAB70923.1"  
 /db\_xref="GI:16550167"  
 /translation="MLITKTAGVFFKSRKRYEFLRSPFHPGTLFLHKIVIGIETSCDDTAAYVDENGVIGEALHSOTVEHLKGGIVPAAOOLHRENIORIVQALSLASGVPSPLSAINATTIKRGLSLGVGSFSLQVGLKKPPIPIHMEAHALITLTKVERPFLVLISGGICLALVQGSDFLLKSLDIANGDMIDKARLPLIKHECEGTMSGGAIIEHAKCNRRFPDIKPLIHAKNCDFPFTGLQHTVDKIINKKECEGTGKQILSSAADIATVQHTWACHLVKRTIRALIFCQKRDLLPQNAVLVASGVASNFCIRRALEILTNATQCTLLCPPRLCTDNGIMIA"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,83e-155 Length: 2208  
 Score: 1944.00 Matches: 386  
 Percent Similarity: 93.24% Conservative: 0  
 Best Local Similarity: 93.24% Mismatches: 4  
 Query Match: 91.48% Indels: 24  
 DB: 9 Gaps: 1  
 US-10-649-273-2 (1-414) x AK055441 (1-2208)  
 Oy 1 MetLeuIleLeuThrIleThrAlaGlyValAlaPhePheLysProSerIleArgIleValTyr 20  
 Db 344 ATGCTAATCTTGATTAAGACGCGAGGATTTTAAACCATCAACAAAGAAAGATTAT 403  
 Oy 21 GlupPheLysrGserPheAnPheHisProGlyThrLeuPheLysIleLysIleValLeu 40

```

Db      404 GAATTTTAAGAAGTTTAATTTTCATCTCGAAGACATATTTCTTCAATAAATAGATTCG 463
Qy      41  G|Y|I|E|G|U|T|H|S|E|C|Y|S|A|P|A|P|H|T|H|A|A|A|V|A|I|A|S|P|G|U|T|H|G|Y|A|S|N|A| 60
Db      464 GGAATTTGAACCTAGTTGTGATGATACAGAGCTGCTGTGTGGATGAATGAACCTGGAATGTG 523
Qy      61  L|E|U|G|Y|G|U|A|I|I|E|H|S|E|C|I|N|T|H|G|U|A|H|S|E|U|Y|S|H|G|Y|G|Y|I|E|V|A|I|P| 80
Db      524 TTGGAGAGAGCAATTCATTCCTCAACCTGAACTTCAATTTAAAAACAGGTGGGATTTCTCT 583
Qy      81  P|R|O|A|A|I|A|G|I|N|G|L|E|U|H|S|A|R|G|U|A|S|N|I|E|G|I|N|A|I|G|I|N|G|U|A|I|E|U|S|E| 100
Db      584 CCAGAGACTCAACAGCTTCAACAGAGAAATATTCACGAAATGACAGAGAGCTCTTCTCT 643
Qy      101 A|A|S|E|R|G|Y|A|I|S|E|P|S|E|S|E|A|P|S|E|U|S|E|A|I|A|E|A|T|H|T|H|I|E|U|S|P|R|O|G|I|N| 120
Db      644 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATTCGCACTAACCTAATAAACAGAGACTT 703
Qy      121 A|A|E|U|S|E|U|G|Y|V|A|I|G|Y|E|U|S|E|P|H|E|S|E|U|G|I|N|E|U|A|I|G|Y|G|I|N|E|U|Y|S|Y|S| 140
Db      704 GCTTTAACTCTGGAGTGGCTTATCATTTACCTTAACAGCTGTAGAGACAGTTAAAGAG 763
Qy      141 P|R|O|H|E|I|P|R|O|I|E|H|S|I|E|W|E|C|U|A|H|S|A|I|E|U|T|H|I|E|A|R|G|E|U|T|H|A|N|Y|S| 160
Db      764 CCATTCATTCCTCATTCATCATATGAGAGGCTCAGCACTTACATTAAGTTGACCAATTA 823
Qy      161 V|A|I|G|U|P|H|E|P|H|E|U|A|I|E|U|E|U|I|E|S|E|R|G|Y|G|Y|H|S|C|Y|E|U|E|U|A|I|E|U|A|I| 180
Db      824 GTAGAAATTTCTTTTATTTAGTTCTTTGATTTCTGAGAGTCACTGCTGTGGCATTAAGTT 883
Qy      181 G|I|N|G|Y|A|I|S|E|A|S|P|H|E|U|E|U|E|U|G|Y|Y|S|S|E|U|A|S|P|I|E|A|I|P|R|O|G|I|N|A|S|P|E|T| 200
Db      884 CAAGAGTTTCAAGATTTCTGCTCTTGGAAAGTCTTTGACATAGCAACAGGTGACAG 943
Qy      201 L|E|U|S|P|Y|S|V|A|I|A|I|A|I|G|U|S|E|U|S|E|U|I|E|Y|H|S|P|R|O|G|I|N|C|Y|S|E|R|H|W|E|S|E| 220
Db      944 CTTGACAGAGTGGCAAGAGACTCTCTTAATTAACCAACAGAGTCTCCACCAAGACT 1003
Qy      221 G|Y|G|Y|V|S|A|I|E|G|U|H|S|E|U|A|I|Y|S|G|I|N|Y|A|N|A|I|G|P|H|S|P|H|E|A|S|P|I|E|Y|S| 240
Db      1004 GGTGGAGAAAGCCATTAAGACATTTGGCCAAACAGAGAAATAGATTTGACATCAACAA 1063
Qy      241 P|R|O|P|L|E|U|H|S|H|S|A|I|Y|S|A|N|C|Y|S|A|S|P|H|E|S|E|P|H|E|T|H|G|Y|L|E|U|G|I|N|H|S|V|A|I|T|H| 260
Db      1064 C|T|C|C|T|T|G|C|A|T|C|A|G|T|T|A|A|A|T|T|G|A|T|T|T|C|T|T|T|A|C|G|A|C|T|T|A|C|A|C|G|T|A|C|T| 1123
Qy      261 A|S|P|Y|S|I|E|I|E|W|E|Y|S|G|U|Y|S|G|U|G|U|G|Y|I|E|G|U|Y|S|G|I|N|I|E|U|S|E| 280
Db      1124 GATTAATTAATTAATGAAGAAAGCAAGAAAGAGAAAGGATTAAGAGAGGGCAAAATCTGCTCT 1183
Qy      281 S|E|R|I|A|A|S|E|P|I|E|A|I|A|I|E|U|H|S|H|T|H|E|A|I|C|Y|H|S|E|U|V|A|I|Y|S|A|I|G| 300
Db      1184 TCACAGACAGACATTTGCTGCACAGTACAGCACAAATGCAATGTCATCTTGTGAAGAA 1243
Qy      301 T|H|I|S|A|T|G|A|I|I|E|U|P|H|E|C|Y|S|Y|S|I|N|A|I|S|P|H|E|U|P|R|O|G|I|N|A|S|N|A|A|I|V|A|I| 320
Db      1244 ACAATCGGGCTATTTCTTTTGTGAAGAGAGACTTTTACTCTCAAAATTAATGACAGTA 1303
Qy      321 L|E|U|V|A|I|A|S|E|R|G|Y|V|A|I|A|S|E|S|A|N|P|H|E|Y|I|E|A|I|G|A|I|A|E|U|G|I|N|I|E|U| 340
Db      1304 CTGGTGTGATCTGGGTGTGCGCAAGTAATCTCTGATCCGCAAGCTCTGGAATTTTA 1363
Qy      341 T|H|E|N|A|I|A|T|H|G|I|N|C|Y|T|H|L|E|U|E|U|C|Y|P|R|O|P|R|O|A|I|G|U|E|U|C|Y|T|H|A|S|P|H|E|N|G|Y| 360
Db      1364 ACAACCGACACAGTGCATTTGTGTGCTCTCTCCAGACTATGACCTATATATGTC 1423
Qy      361 I|L|E|U|T|I|E|A|I|T|P|A|S|N|G|Y|I|E|G|U|A|I|G|U|A|I|G|Y|L|E|U|H|S|A|S|P| 380
Db      1424 ATTATGATTCG----- 1435
Qy      381 I|L|E|G|U|Y|I|E|A|I|G|Y|U|P|R|O|Y|S|C|Y|S|P|R|O|E|U|G|Y|V|A|I|S|E|R|Y|S|G|U|A|I| 400
Db      1436 -----TGATGCTCTTGGAGTGAAGTATCAAAAGAGATT 1471

```

Qy 401 G|Y|G|U|A|I|S|E|I|E|Y|V|A|I|P|R|O|G|I|N|E|U|S|E|U|G|I|N|E|U| 414  
 Db 1472 GGAAGCTTCCATTAATAAGTACCAATTAATAATGACAGTA 1513

RESULT 9  
 BC058172  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus  
 Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Ushed, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullany, S.J., Bosak, S.A., McGwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shvachenko, Y.,  
 Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butlerfield, Y.S., Krzywinski, M.I., Skalek, U., Smailus, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL  
 MEDLINE  
 PUBMED  
 22388257  
 12477932  
 2 (bases 1 to 1844)  
 Strausberg, R.  
 Direct Submission  
 Submitted (15-SEP-2003) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

REMARK  
 COMMENT  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghghi, P.,  
 Hansen, N., Ho, S.-L., Karlins, B., Kong, P., Laric, P., Legaspi, R.,  
 Maduro, Q.L., Masello, C., Maskeri, B., Mastrisano, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Stentrop, S., Thomas, P.J., Touchman, J.W.,  
 Teurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>  
 Series: IRAK Plate: 123 Row: d Column: 22  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 21312463.  
 Location/Qualifiers

```

source
1. .1844
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="MGC:67870 IMAGE:5012054"
/tissue_type="Mammary tumor metastasized to lung. Tumor
arose spontaneously from a senescent normal mammary
(cloneal) outgrowth infected with the virus MTV."
/clone_id="NCI CGAP Lu29"
/lab_host="MDA10B"
/name="Vector: pCMV-SPORT6"
222. 1466
/codon_start=1
/product="Unknown (protein for MGC:67870)"
/protein_id="AAH58172.1"
/db_xref="gi:34849664"
/translation="MLMRLRTAGAIIPKPKSKYGFLLRPSVHPRTLSCKLVLGIET
SCDPTGAAYVDENGNVLGALHSQTOVHLTKGIVPAOOLHRENTOLIVEETLSAS
RITPDSLSAIAITTKRGLALSIGLSPFSQLVNRFKPKPIPIHMHAAHLTRITNK
VEFPLVLILISGHCIALVGVSDPFLIKSLDIAGDMIDKVARLSLTKPESTG
MSGKAIHLAKDGNRFHPTINPMQNAKCDSEFTSLQHTDLITHKKESEITEG
KIISSADIDAAVQHAATACHLAKTRTHAIIFCKKXNLSPANAVLVSGSVASMLYTR
KALIVANATCCTLACPPRLCTDNGIMIANWIERLRAGLGLVDEDIRBPKPL
GIDISREVAEAIKVPRLKML"
333. 1397
/note="Q917: Region: Metal-dependent proteases with
possible chaperone activity [posttranslational
modification, protein turnover, chaperones]"
/db_xref="CDD:COG0533"

misc_feature
333. 1397
/note="Q917: Region: Metal-dependent proteases with
possible chaperone activity [posttranslational
modification, protein turnover, chaperones]"
/db_xref="CDD:COG0533"

ORIGIN

Alignment Scores:
Pred. No.:      2.71e-146      Length:      1844
Score:          1835.00      Matches:      352
Percent Similarity: 91.06%      Conservative: 25
Best Local Similarity: 85.02%      Mismatches:  37
Query Match:    86.35%      Indels:      0
Gaps:           10

US-10-649-273-2 (1-414) x BC058172 (1-1844)

Oy      1  MecLeuIlleuThrMyThAlaGlyValPhePheLysProSerLysArgLysValTyr  20
Db      222 ATGCTAAAGTTAAAGAGACACAGACGACTCTTCCAAAGCCCCCAAGATGAAGTTTAT  281
Oy      21  GluPheLysuArgSerPheAsnPheHisProGlyThrLeuPheLysHisLysIleValLeu  40
Db      282 GGATTTTAAAGAAAGATTGATGTCATCCCAAGAACTCTCTTGTCATAAACTGGCTCTG  341
Oy      41  GlyIleGluThrSerCysAspAspThrIleAlaIleValIleAspGluThrGlyAsnVal  60
Db      342 GGAATTGAAACCACTGTGATGACACAGAGCGCTGTGTGATGAAGAACTGGAAATGTG  401
Oy      61  LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro  80
Db      402 CTGGGGGAAGACCTGCACATCCCAACTCAGGTTTCATGAAAAACAGTGGGATGTTCTCT  461
Oy      81  ProIleAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer  100
Db      462 CCAGTAGCTCAACCACTTCACAGAGAAAATTTCAACCAAAATGATGAAGAAACTCTTCT  521
Oy      101  AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu  120
Db      522 GCAGTAGAATACCAACCAAGCAGATCTTCGGCAATTGCACTACCATCAAAACGGGACTG  581
Oy      121  AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys  140
Db      582 GCCCTAAGCTTAGAGATCGGCTTATCTTTAGCTTACACGCTAGTAATACGGTTAAAAAG  641
Oy      141  ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys  160
Db      642 CGGTTCACTCCGATTCATCATCATGAGGCTCACGACATGATTAATAGGCTCACCAATAAA  701

```

Qy	161	ValGlnpHeProPheLeuValLeuLeu11SeSc1Yc1Yh1sCySLeuLeuA1LeuVal1	180
Db	702	GTAAATTTCTTCTTTCAGTCTTTTAATTTCTGGCGGTCACTGCTGTTGGCATTTAGTC	761
Qy	181	G1nG1yValSerSpPheLeuLeuG1YLeSerLeuAsp11ealaproG1yAspMet	200
Db	762	CAAGTGTTCCTCCATTTCTTGCTCTCTGGGAAGTCTTTGGACATAGCCGAGGCATG	821
Qy	201	LeuAspLySValAlaAaGhGLeuSerLeu11elysh1sProG1uCySeSerThMeSer	220
Db	822	CTTGACAAGGCGGACAGAGAAGCTTTCTTTATCAAAATCCAGATATTTCTACATAGAT	881
Qy	221	GlYc1YLySAla11eG1uH1sLeuAla1ySeG1nG1YAsnArgPheH1sPheAsp11elys	240
Db	882	GGTGAANAAGCTTAAGACATTGTGGCCAGAACGGAATAGATTCCATTTTACTATCAAT	941
Qy	241	ProProLeuH1sH1sAla1yAsnCyAspPheSerPheThrg1yLeuG1nH1sVal1Thr	260
Db	942	CCACTTATMGACAGATGCTAAGAAATTCGATTTTCTTTTACGGGACTTCAACATATTA	1001
Qy	261	AspLyS11e11eMetLybLySg1uLyG1uG1uG1y11eG1uLySg1Yg1n11eLeuSer	280
Db	1002	GATAGAGCTAAATMACACACAGAGAAAAGAAAGAGCATGTGAAGGGGCAATCTGTCA	1061
Qy	281	Ser11a1aAsp11eAla1aThrValG1nH1sThrMeAla1CySh1sLeuVal1ySArg	300
Db	1062	TCAGCTGACAGCATTTGCTGCTGGGTACAGATCCACAGCGTCCACCTTGCAAAAAGA	1121
Qy	301	ThH1sArgAla11eLeuPheCyAluSg1nArgAspLeuProG1nAsnAla1yAla1	320
Db	1122	ACACATCGTGCTATTCGTCTTTTGGACAGAAAATTTGCTATCTCCAGCTTAACGCA	1181
Qy	321	LeuValAlaSerG1yG1yValAlaSerAsnPhyT11eArgArgAla1eG1u11eLeu	340
Db	1182	TTAGTGTATGTGAGGCGTGGCAAGTACTGTACATCCAGAAAAGCATTTGAAATGTCT	1241
Qy	341	ThH1sAla1Thrg1nCySThLeuLeuCySProProAla1yLeuCySThAspAsnG1y	360
Db	1242	GCAATGCAAGCAGCTGCACTTGCTGTCTCCCTCCAAAGCTGTGACATGACATGCG	1301
Qy	361	11eMet11eAla1TPAsnG1Y11eG1uArgLeuArgAlaG1Y1eG1Y11eLeuH1sAsp	380
Db	1302	ATCATGATTCATGGAATGAATGAATGAATTAAGTTCGCTGCGCTTGCTTTATCATGAT	1361
Qy	381	11eG1uG1Y11eArgTyrG1uProLySProLeuG1YValAsp11eSerLyG1uAla1	400
Db	1362	GTAGAGACATCCGATTAAGAACCAAAATGCTCTTGGAATTAAGATATTCAGAGAGT	1421
Qy	401	G1yG1uAlaSer11eLySValProG1nLeuLySMeTGu11e1414	
Db	1422	GCAGAGCTGCCATTAAGATACCGCGATTAATAAATGGACATT1463	
RESULT 10			
LOCUS	ARS41929	1416 bp	DNA linear PAT 08-OCT-2004
DEFINITION	Sequence 177 from patent US 6743619.		
ACCESSION	ARS41929		
VERSION	ARS41929.1	GI:53934009	
KEYWORDS			
SOURCE			
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1416)		
AUTHORS	Tang,Y.T., Zhou,P., Goodrich,R., Liu,C., Asundi,V., Ren,F., Zhang,J., Zhao,Q.A., Yang,Y., Xue,A.J., Wehrman,T., Wang,J.-R., Wang,D., and Drianaac,R.T.		
TITLE	Nucleic acids and polypeptides		
JOURNAL	Patent: US 6743619-A 177 01-JUN-2004;		
FEATURES	Location/Qualifiers		
source	1..1416		
	/organism="unknown"		
	/mol_type="genomic DNA"		

## ORIGIN

## Alignment Scores:

Pred. No.: 5.9e-139 Length: 1416  
 Score: 1747.00 Matches: 340  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 99.71% Mismatches: 0  
 Query Match: 82.21% Indels: 0  
 DB: 6 Gaps: 0

US-10-649-273-2 (1-414) x ARS41929 (1-1416)

QY 74 LysThrGlyGlyValLeuAlaProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArg 93  
 DB 280 AGAACAGGTGGAGATGTTCTTCCACAGCTCAACAGCTTCAACAGAAATATTTCAACCA 339  
 QY 94 IleValGlnGluAlaLeuSerAlaSerGlyValIleSerProSerAspLeuSerAlaIleAla 113  
 DB 340 ATAGTACAGAAAGCTCTTCTCCAGTGGAGTCTCTCCAGTGAAGCTTCAGCAATTGCA 399  
 QY 114 ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133  
 DB 400 ACTACCAATAAACAGAGCTTCTTAAAGCTGGAGTGGCTTATCATTTAGCTTACAG 459  
 QY 134 LeuValGlyGlnLeuLysLeuProPheIleProIleHisIleMetGluAlaHisAlaLeu 153  
 DB 460 CTGGTAGGACAGTAAAGCATTCATCCATTCATCATATGAGGCTCATGACATT 519  
 QY 154 ThrIleArgLeuThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyGly 173  
 DB 520 ACTATTAGTGTGACCAATAAAGTAAATTTCTTTTAAATTTCTTTTAAATTTCTGGAGGT 579  
 QY 174 HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeu 193  
 DB 580 CACTGTCTGTGGCATTTAGTTCAGAGAGTTTCAGATTTCTCTCTTGGAAAGCTTTTG 639  
 QY 194 AspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHis 213  
 DB 640 GACATAGACACAGGAGCATGCTTCAACAGGTGGCAAGAAAGCTTTCTTAATAAAGCAT 699  
 QY 214 ProGluCysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsn 233  
 DB 700 CCAAGAGTCTCCACCATGATGGTGGAGAAAGCCATAGAAATTTGGCCCAACAGCAAT 759  
 QY 234 ArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPhe 253  
 DB 760 AGATTTCAATTTGACATCAAACTCTCTTGCATCAAGCTTGAATAATTTGATTTCTTTT 819  
 QY 254 ThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIle 273  
 DB 820 ACTGAGCTTCAACAGTACTGATTAATAATTAATGAAGAAAGAAAGAGAGGATTT 879  
 QY 274 GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMet 293  
 DB 880 GAGAGAGGGGCAAACTCTCTTCAAGCAGACATGTGCTGCACAGTACAGACACAATG 939  
 QY 294 AlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeu 313  
 DB 940 GCATGCTCACTTGTGAAGAAAGCAACATGGGTATTTCTTTTGAAGAGAGAGCTTG 999  
 QY 314 LeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIle 333  
 DB 1000 TTACCTCAAAATTAATGCACTGCTGATTCATCTGATGCTGCGAAGTAACTTCTATATC 1059  
 QY 334 ArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLysCysProPro 353  
 DB 1060 CGCAGAGCTCTGGAATTTAAACAAAGCAACAGTGCATTTGTGTGCTCTCTCC 1119  
 QY 354 ArgLeuCysThrAspAsnGlyIleMetIleAlaTTPAsnGlyIleGluArgLeuArgAla 373  
 DB 1120 AGATATATCACTGATTAATGACATTAATGATTCATGAATGATTAAGAAAGTACGCTCT 1179  
 QY 374 GlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGly 393

DB 1180 GGCTTGCGCATTTTACATGACATAGAGGATCCGCTATGTAACCAAAATGCTCTTGA 1239  
 QY 394 ValAspIleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGlu 413  
 DB 1240 GTAGACTATCAAAAGAGTTGAGAGAGCTTCCATTAAGTACCAATTAATAATGAG 1299  
 QY 414 Tle 414  
 DB 1300 ATA 1302  
 RESULT 11  
 BC078974  
 LOCUS  
 DEFINITION  
 Rattus norvegicus cDNA clone IMAGE:1111906, partial cds.  
 ACCESSION  
 BC078974  
 VERSION  
 BC078974.1 GI:50926879  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 1546)  
 Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
 Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,  
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heieh,F.,  
 Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
 Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
 Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,  
 Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,  
 Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,  
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Huylk,S.W.,  
 Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
 Fahey,J., Helton,E., Kettelman,M., Madan,A.C., Rodrigues,S.,  
 Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
 Butlerfield,Y.S., Krzywinski,M.I., Skalska,U., Smalys,D.B.,  
 Scherch,A., Schein,J.E., Jones,S.J., and Marra,M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932  
 2 (bases 1 to 1546)  
 DIRECTOR MGC Project.  
 DIRECTOR MGC Project.  
 DIRECTOR MGC Project.  
 Submitted (02-AUG-2004) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Howard Jacobs  
 CDNA Library Preparation: Express Genomics  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-ehgc.stanford.edu>  
 Contact: (Dickson, Mark) [md@paxil.stanford.edu](mailto:md@paxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
 R. M.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Series: tRAK Plate: 192 Row: F Column: 5  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Hexamer frequency ORF  
 analysis, Similarity but not identity to protein.  
 Location/Qualifiers  
 FEATURES

source

1. 1546  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
/db\_xref="IMAGE:7111906"  
/clone\_lib="NIH MGC\_237"  
/lab\_host="DH10B"  
/note="vector: pBmpres1"  
144. .>1546  
/codon\_start=1  
/product="unknown (protein for IMAGE:7111906)"  
/protein\_id="AAH78974.1"  
/db\_xref="GI:50926880"  
/translation="MLMSKTRNGATPPRRSVNRGIRRFNVQPALFHHKIVLGIET  
SCDTPAAVVDENGLVEALSHSOTEVHUKTGIVPVAQQLRENIQIVAEALSAS  
GVSPSDSAIAITTKPGALISVGLSFSVOLVNPQKPIPIHMEALALIRLTHK  
VGFPLVLLISGGILALVQSDPLGSLDIPAGMDLKVARRSLIKHPCST  
MSGKAIELHAKENRFPETINPWNANKCPSPFGLOHDTKLITHEKEGIEKG  
QLSSADIAAIVOHATACHLAKRTHRALIPCOQNLSPAAVIVSGVANSYIR  
RALEIVANATCTLCPPRCTDNGIMIANNGIRLRAGLGLIDVEDIRPREPAGE  
IVSWLRVLTALTEDLDVSSTHTVAHSPUNSGSRANIQTSKCSCTQIVMKIVRHIL  
NINERSKKKKKKKK"

ORIGIN

Alignment Scores:  
Pred. No.: 4.9e-137 Length: 1546  
Score: 1725.00 Matches: 335  
Percent Similarity: 91.00% Conservative: 19  
Best Local Similarity: 86.12% Mismatches: 35  
Query Match: 81.18% Indels: 0  
DB: 10 Gaps: 0

US-10-649-273-2 (1-414) x BC078974 (1-1546)

QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
DB 144 ATGCTTAATGTTAGTAAAGACAGACAGAGCTATTCGAGACACCAAGAGTAACTGCT 203  
QY 21 GluPheLeuArgSerPheAspPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
DB 204 GGATTATTAAGAAATTAATGTTCACTAGACCACTTTTCATCATAACTGGTGGCTG 263  
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGlyAsnVal 60  
DB 264 GGAATTTGAACACAGCTGTATGACACAGACGCTGCTGTGATGATAAATGGAATATG 323  
QY 61 LeuGlyGluAlaIleHisSerGlnPheGlyValHisLeuLysThrGlyIleValPro 80  
DB 324 CTGGAGAGAGCACTGCATCCAGACGGAAGTCCATCTGAAAAACAGGTGGGATTGTCT 383  
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
DB 384 CCAAGTCGCTCAACAGCTTCAAGAGAAATATTCACGCAATAGTAAAGAGCTTTCT 443  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
DB 444 GCCAGTGGGGTCTCCCAAGCAGACCTCTCAGCAATTCACATCAACACAGAGACTG 503  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysIle 140  
DB 504 GCCCTTAAGCTGGGGTGGCTTATCTTTAGCGTACAGTAAATCACTTTTAAAG 563  
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
DB 564 CCAATTATCCCAATCATATGAGAGCTCAGCGCTGACTTATAGGCTGACCCACAA 623  
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
DB 624 GTCCGATTTCTTTTATGTTCTTTGATTTCGAGAGCCACCTGCGTGGGCTTATG 683  
QY 181 GlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200

DB 684 CAGAGTGTTCAGATTCTCTCTCGGAGAACTCCCTGAGCATACGCGCAGACATG 743  
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220  
DB 744 CTTGACAGAGTGGCAAGAGACTTTCTTATATCAATCCAGAAATGTTCTACAAATGAGT 803  
QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
DB 804 GGGGGAAGACTATACAACTTGGCCAAAGAAAGAAATGATTCACCTTACTATCAAT 863  
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB 864 CCAACCATGAGATCTAAGAACTGTGATTTCTTTTACGGAACCTTCAACATGTCACC 923  
QY 261 AspLysIleIleMetLysLysGlyLysGlyGluGlyIleGluLysGlyGlnIleLeuSer 280  
DB 924 GATTAACCTAATTAACACAGAGAAAGAAAGAGGATTAAGAGGGGCAAACTCTTCA 983  
QY 281 SerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
DB 984 TACGCCGACAGACTTCTGCTGCGGTACAGCAGCAACGCGTCCACTTGGGAAAGA 1043  
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAlaVal 320  
DB 1044 ACACATGCTGCTATTCCTGTTTCCAGCAAGAAATTTGCTATCTCCAGCTAACGCACTA 1103  
QY 321 LeuValAlaSerGlyGlyValAlaSerAspPheThrIleArgArgAlaLeuGluIleLeu 340  
DB 1104 TTAGTGTGCTCGAGAGTGTTCGAAATCTTATACATCCGAAAGAGATTGAAATGTGA 1163  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgArgLeuCysThrAspAsnGly 360  
DB 1164 GCAAATGCACACAAATGCACCTTGTGTGTGCCCCCTCCAGACTGTGCACATAGT 1223  
QY 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
DB 1224 ATCATGATTGACATGATGAATGAATTAAGATTAAGCTGCTGGCTGGCAATTTACATGAT 1283  
QY 381 IleGluGlyIleArgGlyArgLysProLys 389  
DB 1284 GTAGAAAGACATCCGATACGAACCAAG 1310

RESULT 12

AR428809 1526 bp DNA linear PAT 18-DEC-2003  
LOCUS AR428809  
DEFINITION Sequence 23 from patent US 6642041.  
ACCESSION AR428809  
VERSION AR428809.1 GI:40188595  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1526)  
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.  
TITLE Polynucleotides encoding a novel metalloprotease, Mp-1  
JOURNAL Patent: US 6642041-A 23 04-NOV-2003;  
FEATURES  
Location/Qualifiers  
1. 1526  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Alignment Scores:  
Pred. No.: 3.95e-108 Length: 1526  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 65.18% Indels: 0  
DB: 6 Gaps: 0

US-10-649-273-2 (1-414) x AR428809 (1-1526)

QY 148 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 167

|||||  
Db 1 ATGGAGGCTCATGCACTTACCTTAAGTGGTACCACCAATTAAGTAATTTCTTTTAAAGTT 60  
Oy 168 LeuLeu1LeSerG1yG1yH1sCySeuLeu1Leu1ValG1nG1yValSerAepPheLeu 187  
Db 61 CTTTGAATTTTCGGAGGTCAGTGTCTGTTGGCATTAGTTCAAGAGCTTTTCAAGTTTCTG 120  
Oy 188 LeuLeuG1yLysSerLeu1Leu1ValA1aProG1yAaspMetLeu1Leu1ValA1aArg 207  
Db 121 CTTCTTGGAAAGTCTTTTGGACATAGACCAAGGTCATCTTGAACAAGTGGCAAGAAG 180  
Oy 208 LeuSerLeu1Leu1Leu1Leu1Leu1Leu1Leu1Leu1Leu1Leu1Leu1Leu1Leu1 227  
Db 181 CTTCTTTTAATTAACATTCACAGAGTGTCTCACATGAGTGGGGAACCAATGAACAT 240  
Oy 228 Leu1a1aLysG1nG1yAaspPheH1sPheAap1Leu1Leu1ProPheH1sH1a1aLys 247  
Db 241 TTGGCAAAACAAGAAATGATTCTTTCATTTGACATCAACCTCCCTTGACATCATGCTAA 300  
Oy 248 AaspCyAaspPheSerPheThrG1yLeuG1nH1sValThAaspLys1Leu1MetLysLys 267  
Db 301 AATTGTGATTTTCTTTTACTGGACTTCAACAGCTTATGATTAATTAATGAAGAAG 360  
Oy 268 G1uLysG1uG1uG1y1LeG1uLysG1n1LeuSerSer1a1aAasp1Leu1a1a 287  
Db 361 GAAAAAGAGAGGATTAAGAGAGGCAATCTGTCTTCAAGCAGACATTCCTGCC 420  
Oy 288 ThrValG1nH1sThMetA1aCyH1sLeuValLysArgThH1sArg1a1LeuPhe 307  
Db 421 ACAGTACAGCACACATGGCATGTCTCTTGAAAAGAACATCGGGCTATTCTGTT 480  
Oy 308 CysLysG1nAArgAaspLeuLeuProG1nAaspA1aValLeuValA1aSerG1yG1y 327  
Db 481 TGTAAAGAGAGAGACTTGTACTTCAAAATATGACATCTGTTGCATCTGTTGTC 540  
Oy 328 A1aSerAepPheTyr1LeArg1a1aLeuG1u1LeuThraSn1aThrg1nCyStr 347  
Db 541 GCAAGTAACTTCTAATATCGGACAGCTCGAAATTTTAAACAACGCAACAGTGCAC 600  
Oy 348 LeuLeuCyProProPheArgLeuCyStrAaspA1a1Met1Leu1a1aThraPang1y 367  
Db 601 TTGTGTCTCTCCCTCCAGACTATGACATGATTAATGATTAATGATTAATGATTAAT 660  
Oy 368 1LeG1a1aArgLeuArg1a1aG1yLeuG1y1LeuH1sAap1LeG1uG1y1LeArg1yG1n 387  
Db 661 ATTGAATACTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
Oy 388 ProLysCyProLeuG1yValAasp1LeSer1yG1uValG1yG1uA1aSer1LeuVal 407  
Db 721 CCAAAATGCTCTTGGAGTAGACATATCAAAAGAAGTTGAGAGCTTCCATTAAGA 780  
Oy 408 ProG1nLeuLysMetG1u1Le 414  
Db 781 CCACAAATTAATAATGAGATA 801  
RESULT 13  
BX934991 1522 bp mRNA linear VMT 02-FEB-2004  
LOCUS BX934991  
DEFINITION Gallus gallus finished cDNA, clone CHEST189114.  
ACCESSION BX934991  
VERSION BX934991.1 GI:41635519  
KEYWORDS  
SOURCE  
ORGANISM Gallus gallus (chicken)  
Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus;  
1 (bases 1 to 1522)  
Boardman, F.E., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.E.,  
Croning, M.U.R., Davies, R.M., Francis, M.D., Grafton, D.V.,  
Hubbard, S.J., Humphray, S.J., Hunt, P.J., Maddison, M., McLaren, S.R.,  
Nblett, D., Overton, I.M., Rogers, J., Scott, C.B., Taylor, R.G.,  
Tickle, C. and Wilson, S.A.

TITLE Direct Submission  
JOURNAL Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: chicken@ms.umbist.ac.uk  
COMMENT BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA  
sequencing project.  
This sequence is from the  
BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,  
from a library constructed by Elizabeth Bosch. cDNA was prepared  
from RNA extracted from muscle, normalised, and poly A-tailed.  
ECORI-NotI cut cDNA was then ligated into the vector. Vector:  
pBluescript II KS(+); Site\_1: EcoRI; Site\_2: NotI Host: Escherichia  
coli DH10B.  
FEATURES  
Location/Qualifiers  
source 1..1522  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="layer and broiler"  
/db\_xref="taxon:9031"  
/clone="CHEST189114"  
/clone\_lib="CSEORBN11"  
/dev\_stage="adult"  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,566-106 Length: 1522  
Score: 1362.00 Matches: 258  
Percent Similarity: 83.20% Conservative: 54  
Best Local Similarity: 68.80% Mismatches: 63  
Query Match: 64.09% Indels: 0  
DB: 5 Gaps: 0  
US-10-649-273-2 (1-414) x BX934991 (1-1522)  
Oy 37 Lys1Leu1ValLeu1G1y1LeG1uThrSerCyAaspPheThra1a1aValA1aSerG1u 56  
Db 254 AGACTGTGTCTGGGCAATCGAAACCACTGTGACACACGGCGGGGTGCTGAGAG 313  
Oy 57 ThG1yAasnValLeuG1yG1uA1a1LeH1sSerG1nThrg1uValH1sLeuLysThG1y 76  
Db 314 GCGGGACAGGTGTGGGAGAGGCGCTGCAGAGCCAAAGAGTCCCTCAAGACAGT 373  
Oy 77 G1y1Leu1a1aProPheA1a1aG1nG1nLeuH1sArgG1uAasn1LeG1a1aValG1n 96  
Db 374 GGAATTAATCTCATGATGACAGACAGCTTCAAGAAAGCATCCAGAAAGTAAAG 433  
Oy 97 G1uA1aLeuSerA1aSerG1yValSerProSerAaspLeuSerA1a1a1aThr1Le 116  
Db 434 GAAGCACTCACTGCGCAGTGAAGTTCTGTAATGAACCTTCTGCTATTGCAACTACAGT 493  
Oy 117 LysPProG1yLeuA1aLeuSerLeuG1yValG1yLeuSerPheSerLeuG1nLeuValG1y 136  
Db 494 AAACCAAGACTTCCGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 553  
Oy 137 G1nLeuLysLeuPhePhe1Leu1Leu1Leu1Leu1Leu1Leu1Leu1Leu1Leu1Leu1 156  
Db 554 AGCTACCAAGAGCTTTCATTAACCATTCATCAATGAGAGCTACAGCATTAATCA 613  
Oy 614 CTGACAGACCAAGTAAATATTCCTTCTTATGTTCTTTACTCTCCGAGAGCTACATG 673  
Oy 177 LeuA1aLeuValG1nG1yValSerAaspPheLeu1Leu1G1yLysSerLeuAap1Leu1 196  
Db 674 TTGGCAAGAGAGAGTGAATTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 733  
Oy 197 ProG1yAaspMetLeuAaspLysValA1aArgArgLeuSerLeu1Leu1Leu1Leu1Leu1 216  
Db 734 CCAAGTACATCTGATTAAGTAAAGTAAAGCTCTTTTAAAGTAAAGCTCCGAGAGTGC 793  
Oy 217 SerThMetSerG1yG1yValA1a1LeG1uH1sLeuA1aLysG1nG1yAaspPheH1s 236  
Db 794 CACGCGATGCTGGGGGAAGGCAATGAGCACTGCTCAACCGGAGCTGGCAAG 853



Qy 237 Pheasp1lelyspProleuHish1salysaenCyaspPheSerPheThrglyleu 256  
 Db 854 TAGCTTTCAGACTTCCATGCAACAGATGTAAGTATTTCTTTCTCCGAGACT 913  
 Qy 257 GlnH1salThraspLygl1leleuThrysglygl1leuThrygl1leuThrygl1leu 276  
 Db 914 CAGAGCTTTCAGACTTCCATGCAACAGATGTAAGTATTTCTTTCTCCGAGACT 973  
 Qy 277 GlnH1leuSerSerAla1alasp1leuThrygl1leuThrygl1leuThrygl1leu 296  
 Db 974 GAAATCTTTCAGACTTCCATGCAACAGATGTAAGTATTTCTTTCTCCGAGACT 1033  
 Qy 297 LeuVallyspThraspLygl1leuThrygl1leuThrygl1leuThrygl1leu 316  
 Db 1034 ATTATTCAGACTTCCATGCAACAGATGTAAGTATTTCTTTCTCCGAGACT 1093  
 Qy 317 AsnAspAlaValleuVal1alasp1leuThrygl1leuThrygl1leuThrygl1leu 336  
 Db 1094 AAACTGCACTTTCAGACTTCCATGCAACAGATGTAAGTATTTCTTTCTCCGAGACT 1153  
 Qy 337 LeuGlu1leuThraspLygl1leuThrygl1leuThrygl1leuThrygl1leu 356  
 Db 1154 CTGCACTTTCAGACTTCCATGCAACAGATGTAAGTATTTCTTTCTCCGAGACT 1213  
 Qy 357 ThraspAspGly1leuThrygl1leuThrygl1leuThrygl1leuThrygl1leu 376  
 Db 1214 ACCGATTCAGACTTCCATGCAACAGATGTAAGTATTTCTTTCTCCGAGACT 1273  
 Qy 377 IleuH1salp1leuThrygl1leuThrygl1leuThrygl1leuThrygl1leu 396  
 Db 1274 ATTATTCAGACTTCCATGCAACAGATGTAAGTATTTCTTTCTCCGAGACT 1333  
 Qy 397 SerlysglyVal1alasp1leuThrygl1leuThrygl1leuThrygl1leu 411  
 Db 1334 TCCTGCACTTTCAGACTTCCATGCAACAGATGTAAGTATTTCTTTCTCCGAGACT 1378  
 RESULT 14  
 LOCUS BC038910 1017 bp mRNA linear ROD 21-OCT-2003  
 DEFINITION Mus musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA  
 clone IMAGE:5053559), partial cds.  
 ACCESSION BC038910 GI:24433548  
 VERSION BC038910.1 GI:24433548  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1017)  
 Strausberg, R., Feingold, E. A., Grouse, L. H., Derge, J. G.,  
 Klausner, R. D., Collins, F. S., Wagner, L., Shenmen, C. M., Schuler, G. D.,  
 Altschul, S. F., Zeeberg, B., Buetow, K. H., Schaefer, C. F., Bhat, N. K.,  
 Hopkins, R. F., Jordan, H., Moore, T., Max, S. I., Wang, D., Hsieh, F.,  
 Ditchenko, L., Marusina, K., Farmer, A. A., Rubin, G. M., Hong, L.,  
 Stepien, M., Soares, M. B., Bonaldi, A. F., Casavant, T. L.,  
 Scheer, T. E., Brownstein, M. J., Usdin, T. B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Rana, S. S., Loquellano, N. A., Peters, G. J.,  
 Abramson, R. D., Mullaly, S. J., Bosak, S. A., McEwan, P. J.,  
 McKernan, K. J., Malek, J. A., Gunaratne, P. H., Richards, S.,  
 Worley, K. C., Hale, S., Garcia, A. M., Gay, L. J., Hu, X., Gibbs, R. A.,  
 Villalon, D. K., Muzny, D. M., Sodergren, E. J., Lu, X., Gibbs, R. A.,  
 Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodigues, S.,  
 Sanchez, A., Whiting, M., Madan, A., Young, A. C., Shevchenko, Y.,  
 Bouffard, G. G., Blakeley, R. W., Touchman, J. W., Green, E. D.,  
 Dickson, M. C., Rodriguez, A. C., Grimwood, J., Schmutz, J., Myers, R. M.,  
 Butlerfield, Y. S., Krzywinski, M. I., Skalska, U., Smalins, D. E.,  
 Schermer, A., Schein, J. E., Jones, S. J., and Marra, M. A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 JOURNAL MEDLINE  
 PUBLISHED 12477932  
 REFERENCE 2 (bases 1 to 1017)

AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-OCT-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
 BC Cancer Agency, Vancouver, BC, Canada  
 info@bcgsc.bc.ca  
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
 Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeay, Steven  
 Ness, Pawan Pandoh, Anna-Liisa Prabh, Parvaneh Saeedi, Jacqueline  
 Schein, Duane Smalins, Michael Smith, Lorraine Spence, Jeff Stott,  
 Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,  
 George Yang, Scott Zuyderduyn, Marco Marra.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Series: IRAX Plate: 86 Row: f Column: 12.  
 FEATURES  
 source  
 1. 1017  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5053559"  
 /tissue\_type="Liver, normal, 5 month old male mouse."  
 /clone\_lib="NCI CGAP\_L19"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 7.91e-99 Length: 1017  
 Score: 1273.00 Matches: 241  
 Percent Similarity: 92.73% Conservative: 14  
 Best Local Similarity: 87.64% Mismatches: 20  
 Query Match: 59.91% Indels: 0  
 DB: 10 Gaps: 0  
 US-10-649-273-2 (1-414) x BC038910 (1-1017)  
 Qy 140 LyPProPhe1leProlleH1sh1smetGluAlaH1salThrygl1leuThrasp 159  
 Db 2 AACCCATTCATCCCGATTCATCAGTGAAGCTTCCAGCTATTTAGCTACCAAT 61  
 Qy 160 LyPValGluPheProlleH1salThrygl1leuThrygl1leuThrygl1leu 179  
 Db 62 AAAGTGAATTTCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 121  
 Qy 180 ValGlnGlyValSerAspPheLeuLeuGlyLySserLeuAsp1leuAlaProGlyAsp 199  
 Db 122 GTCCAGAGTGTTCCTCCATTTCTGCTCTGGAACTTTGGACATAGCAGAGCGAC 181  
 Qy 200 MetLeuAspLyValAlaValGluLeuSerLeuLeuLySsh1spProGlyCysSerThmet 219  
 Db 182 ATCTGTGCAAGGAGGCGAAGAGCTTTTATTCACAACTCCAGATGTTCTACATG 241  
 Qy 220 SerGlyGlyValAla1leuGlyH1salThrygl1leuThrygl1leuThrygl1leu 239  
 Db 242 AGTGTGGAAGGATTAAGACACTTGGCAAGAGGCAAGATTAATTTCAATTTACTATC 301  
 Qy 240 LyPProPheH1sh1salThrygl1leuThrygl1leuThrygl1leuThrygl1leu 259  
 Db 302 AATCCACTTATGCAAGATGCTAAGATTTGATTTCTTTCAAGGAGCTTCAATATTT 361



Oy		260	thrtaapysllellemetylsgsluyluglylllegluylsgtynlleleu	279
Db		362	actgataagctaatatcacacacaagaamaaaaggcattggagaaaatcccg	421
Oy		280	serseatalaalasplleallalathrvlaglnhsthmetalcyshtisleuvalvs	299
Db		422	tcatcagctgcagacatttcctctcggtaccgatgcacacagccttgccaccttgcgana	481
Oy		300	argtnrhlsargalaalleubhecyslugsclnarspleuleuproglinahsanala	319
Db		482	agAACACATCGGCTATTCTGTGGTTTGGACAAGAAAAAATTCCTCCTCCAGCTAAACGA	541
Oy		320	valleuvallaaserglylvalalaserasnphetyrlsarqargalaleugluile	339
Db		542	gtatttagttgmatctcggaagcttccaagttaaCTTGATCATCCGAAAAGCATTGGAAAAAT	601
Oy		340	leuthranalathrglncysrthrleucycsproporproargleucyrthrapsen	359
Db		602	gtcgcaaaatgcacacgcacgcacgtgtgtgtctccacctcCAAGACTGCGACTACAT	661
Oy		360	glyllemeleialatrpsansglylleglualsuarqalaglVleuglylleuhis	379
Db		662	ggcatcatcattgcattgcattggaattgaaagaattacgtctgcgttgcccttttacct	721
Oy		380	aspillegluglylleargtryglubprolysCysProleuglyvalasplleserlgsu	399
Db		722	GATGTAAGAAGACATCGATATGACCAAAATGCTCTCTTGAGTAGACATATCCAGAGA	781
Oy		400	valglyglualaaserllelysvValProglneulysmetgluile	414
Db		782	gttgacagaagctgccattaaagttacccgatttAAAAAATGCACCT	826

  

RESULT 15

LOCUS	BX930963	1558 bp	mRNA	linear	VFT 30-MAR-2004
DEFINITION	Gallus gallus finished cDNA, clone CHEST62n16.				
ACCESSION	BX930963				
VERSION	BX930963.2 GI:46016890				
KEYWORDS					
SOURCE	ORGANISM				
	Gallus gallus (chicken)				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Archosauromia; Aves; Neognathae; Galliformes; Phasianidae;				
	Phasianine; Gallus.				
	1 (bases 1 to 1558)				
REFERENCE	Boardman,P.E., Bonfield,J.K., Brown,W.R.A., Carder,C., Chalk,S.E.,				
AUTHORS	Croning,M.D.R., Davies,R.M., Francis,M.D., Graham,D.V.,				
	Hubbard,S.J., Humphray,S.J., Hunt,P.J., Maddison,M., McLaren,S.R.,				
	Niblett,D., Overton,I.M., Rogers,J., Scott,C.E., Taylor,R.G.,				
	Tickle,C. and Wilson,S.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (29-Mar-2004) Sanger Institute, Hinxton, Cambridgeshire,				
	CB10 1SA, UK. E-mail enquiries: chickest@hms.unist.ac.uk				
COMMENT	On Apr 1, 2004 this sequence version replaced gi:41631491.				
	BBSRC/Dundee/Nottingham/Sanger/Sheffield/UWIST Gallus gallus cDNA				
	sequencing project.				
	This sequence is from the				
	BBSRC/Dundee/Nottingham/Sanger/Sheffield/UWIST cDNA collection,				
	from a library constructed by Elizabeth Bosch. cDNA was prepared				
	from RNA extracted from limbs,				
	and poly A-titrmed. EcoRI-NotI cut cDNA was then ligated into the				
	vector. Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI				
	Host: Escherichia coli DH10B.				
	Location/Qualifiers				
	1..1558				
FEATURES	/organism="Gallus gallus"				
source	/mol_type="mRNA"				
	/strain="White Leghorn, Hisex"				
	/db_xref="taxon:9031"				
	/clone="CHEST62n16"				
	/clone_lib="CSEQCNU20"				
	/dev_stage="stage 36"				

ORIGIN	Alignment Scores:	Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:
US-10-649-273-2 (1-414) x BX930963 (1-1558)	4.57e-93	1208.00	82.54%	67.46%	56.85%	5	
	Length: 1558	Matches: 228	Conservative: 51	Mismatches: 59	Indels: 0	Gaps: 0	
QY	74	LYSTHrGlyGlyIleValProProAlaIaIaGInGInLeuHiaArgGInaenIleGInArg	93	.....	.....	.....	
Db	486	AGAGCAGGTGGAAATATCTCTATGTAGCAGACAGCTTCAAGAAAGACATCCAGCA	545	.....	.....	.....	
QY	94	IleValGInGInaIaleuSerAlaSerGlyValaSerProSerAAspIeuSerAlaIleAla	113	.....	.....	.....	
Db	546	GTAGTAAAGGAAGCACTCAGTCCAGTGGAGTTTCTGTAATGAATCTTCTGTATTTGCA	605	.....	.....	.....	
QY	114	ThrThrIleLysProGlyLeuAlaIeuSerLeuGlyValaGlyLeuSerPheSerLeuGIn	133	.....	.....	.....	
Db	606	ACTACAGTGAACCAAGCACTTGCATGACCTTGAGAGTGGAGCTGCAGTACAGCTTACAG	665	.....	.....	.....	
QY	134	LeuValaGlyGInLeuLysLysProPheIleProIleHiaMetGInaIaHiaIaLeu	153	.....	.....	.....	
Db	666	CTGGTGAAGGGTACAGAAAGCTTTTCATACCATTCAACATGAGAGCTCAGCACTT	725	.....	.....	.....	
QY	154	ThrIleArgLeuThrAAsnLysValaGluPheProPheLeuValaIeuLeuIleSerGlyGly	173	.....	.....	.....	
Db	726	AGCAACAGACTGACAGAGCCAGTAAATTTCCCTTTAGTTCTTTACTCTCCGAGGT	785	.....	.....	.....	
QY	174	HisCysLeuLeuAlaIeuValaGInGlyValaSerAAspPheLeuLeuLeuGlyLysSerLeu	193	.....	.....	.....	
Db	786	CACATCATCTTGGCAGTAGCAGAAAGAGTTTCAATTTCTTCTGCTTGAACAGTCCATA	845	.....	.....	.....	
QY	194	AspIleAlaProGlyAAspMetLeuAspLysValaIaArgArgLeuSerLeuIleLysHis	213	.....	.....	.....	
Db	846	GATATAGACCCAGGGTACATGCTGTAAAGTAAAGGAAGAGCTCTTTAGTGAAGCAC	905	.....	.....	.....	
QY	214	ProGInCysSerThrMetSerGlyGlyLysAlaIleGInHiaIeuValaLysGInGlyAAsn	233	.....	.....	.....	
Db	906	CCGAGTCCACAGGATGGCTGGGGGAGGAAAGGAAATAGACACTGGCTTCAACCGAGAC	965	.....	.....	.....	
QY	234	ArgPheHiaPheAspIleLysProProLeuHiaHisAlaLysAAsnCysAAspPheSerPhe	253	.....	.....	.....	
Db	966	TGGCAACAGTACACTTTCAGACTTCCCAAGCAACAGTACGTAATGTAATTTTCTTTC	102	.....	.....	.....	
QY	254	ThrGlyLeuGInHiaIleValaThrAspLysIleIleMetLysLysGInLysGInGlyIle	273	.....	.....	.....	
Db	1026	TCCGACATTCAGAGGCTTGTCAAAAGCCATTCTTCGAAAGAAAGAAAGAAAGAGTATT	108	.....	.....	.....	
QY	274	GluLysGlyGInIleLeuSerSerAlaIaAspIleAlaIaThrValaGInHiaThrMet	293	.....	.....	.....	
Db	1086	CAGAAAGGGAATCTCTGCTCCGTTAAGACATTCGCTGCTGCACAGCATGATAGTG	114	.....	.....	.....	
QY	294	AlaCysHiaIeuValaLysArgThrHiaArgAlaIleLeuPheCysLysGInArgAspLeu	313	.....	.....	.....	
Db	1146	GCTGCTCATATATATCCAGGGAGCACACGAGCCATGCTCTTGTGATGAAAAACAGATA	120	.....	.....	.....	
QY	314	LeuProGInaAsnAlaValaLeuValaIaSerGlyGlyValaIaAsenPhePheTyIle	333	.....	.....	.....	
Db	1206	TTATTTACCAAAAACAGCACTCTGCTGTATATGAGAGGGTTCGAAATATCATGATATTC	126	.....	.....	.....	
QY	334	ArgArgAlaLeuGInIleLeuThrAAsnAlaThrGInCysThrLeuLeuSerProProPhe	353	.....	.....	.....	
Db	1286	AGAAAGAGCATGCAACCTCTGGCAAAATGAAACGTTTGTCTTTTCTGCTCTCTCCA	132	.....	.....	.....	
QY	354	ArgLeuSerThrAAspAsnGlyIleMetIleAlaThrPAsnGlyIleGluArgLeuArgAla	373	.....	.....	.....	
Db	1326	AGCGCTGTCCACGAAATATGCTTTATGATTCATGGAATGGCAATTCGAAAGGTTGCTGTCA	138	.....	.....	.....	
QY	374	GlyLeuGlyIleLeuHiaAspIleGluGlyIleArgTyGInLysProLysCysAAspLeuGly	393	.....	.....	.....	

```

Db      1386  GGATGTGTATTTATACAGTACTGATGCGATCCGCTATGAACCAAGCTCCCTTGA 1445
Qy      394   ValAspIleSerIyGluValGlyGluAlaSerIleIyValProGlnLeuLys 411
      1446  ATTGATATTTCCAAAGAGTGAAGAGGATTCCATCAGAGTGCCAAGACTAAG 1499

```

Search completed: November 10, 2005, 22:36:40  
 Job time : 8193.05 secs



PI Mao Y, Xie Y;  
XX MPI: 2002-115090/16.  
DR P-PSDB; ABB05481.

XX O-sialoglycoproteinase-like protein and encoding polynucleotide, useful  
PT for diagnosing, preventing and treating related diseases.

XX Claim 5; Page 29-30 (Disclosure); 38pp; Chinese.

XX The present sequence encodes human O-sialoglycoproteinase-like protein  
CC (OSGPLP). The present invention also describes: (1) the preparation of  
CC the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the  
CC prevention and/or treatment of related diseases; (4) utilizing the OSGPLP  
CC protein in screening its agonist, excitomotor and inhibitor and preparing  
CC an antibody against the OSGPLP protein; and (5) the use of the OSGPLP  
CC polynucleotide sequences, proteins, agonists, excitomotors, inhibitors  
CC and antibodies in treating diseases related to the abnormal OSGPLP gene  
CC and in preparing the medicine composite for the treatment

XX Sequence 2058 BP, 637 A; 400 C; 410 G; 611 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	1,94e-212	Length:	2058
Score:	2125.00	Matches:	414
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2 (1-414) x ABA93268 (1-2058)

QY 1 MetLeuileLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValIyr 20  
Db 110 ATGCTAATCTTGACATTAAGAGCTCAGAGAGTTTATTTTAAACCATCAAAAGAAAGTTTAT 169  
QY 21 GluPheLeuArgSerPheAsnPhenHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
Db 170 GAATTTTAAAGAAAGTTTAAATTTTCATCTGGAACCTATTTCTCATATAAAAGATATG 229  
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaValAspGluThrGlyAsnVal 60  
Db 230 GGAATTTAAACTAGTTGATGATATACAGAGAGCTGCTGTGTGATGAATGAACTGGAAATG 289  
QY 61 LeuGlyValAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80  
Db 290 TTGGAGAAAGCAATACATTTCCCAACTGAATTCATTTAAAAACAGTGGATTTGTTCT 349  
QY 81 ProAlaAlaGlnGlnLeuHisArgGluValenIleGlnArgIleValGlnGlnAlaLeuSer 100  
Db 350 CCAGCAGCTCAACACCTTCACAGAGAAATATTCACAGAAATATACAGAAAGCTCTTTCT 409  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
Db 410 GCCAGTGGAGTCTCTCAAGTACCTCTCAGCAATTCGAACTACCTAATAACAGAGACTT 469  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140  
Db 470 GCTTTAAGCTCGAGAGTGGCTTATCATTTAGCTTACAGTGTAGAGACGTTAAAAAG 529  
QY 141 ProPheIleProIleHisIleMetGluValHisAlaLeuThrIleArgLeuThrAsnLys 160  
Db 530 CCATTCATTCATTCATCATATATGAGGCTCATGCACTTATTTAGTTGACCAATTA 589  
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
Db 590 GTAGAAATTTCTTTTATGTTCTTTTGAATTTCTGAGAGTACATGCTGTGGCAATTA 649  
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyLysPhe 200  
Db 650 CAAGAGTTTCAATTTCTGCTTCTTGGAAAGTCTTTTGACATAGCACAGATGACATG 709  
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220

Db 710 CTGACAAAGTGGCAAGAGACTTTCTTTAATAAAACATCCAGAGTCTCCACATGAGT 769  
QY 221 GlyIleLysValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
Db 770 GGTGGGAAAGCCATTAACATTTTGGCCAAACAGAAATAGATTTTCAATTTTGACATCA 829  
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
Db 830 CTTCCCTTGCAATCATCTTAAATTTGTATTTTCTTTTAAGTGAAGCTTCAACAGCTTACT 889  
QY 261 AspLysIleIleLysLysLysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSer 280  
Db 890 GATTAATATATATGAAAAAGGAAAAAGGAAAGTATGAGAAAGGCAAAATCTGTCT 949  
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLysLeuValLysArg 300  
Db 950 TCAGCAGACAGATTCCTCCACAGATACAGACACAAATGCAATGCTATCTTGTGAAGA 1009  
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320  
Db 1010 ACAATCGGAGCTATTTCTGTTTGTATAGCAGAGAGACTTGTTACTCAAAATATGACGTA 1069  
QY 321 LeuValAlaSerGlyValAlaLysAsnPhenIleArgArgAlaLeuGluIleLeu 340  
Db 1070 CTGTTGCAATCTGTGTGTGTCGCAAGTAATCTTATATCCGACAGAGCTCGAAATTTTA 1129  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgArgLeuCysThrAspAsnGly 360  
Db 1130 ACAACGCAACACAGGCACTTGTGTGTGCTCTCCACGACTATGCACTGTAAAGGC 1189  
QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
Db 1190 ATTATATTCAGAGAAATGATTAAGACTACGAGCTGCTTGAGCAATTTTACATGAC 1249  
QY 381 IleGluGlyIleArgGlyArgLysCysProLeuGlyValAspIleSerLysGluVal 400  
Db 1250 ATGAAAGCATCCGCTATGACCAAAATGTCCTTGGAATGACATATCAAAAGAGTT 1309  
QY 401 GlyIleLysSerIleLysValProGlnLeuLysMetGluIle 414  
Db 1310 GGAGAAAGCTTCATTAAGTACCAATTAATAATGAGATA 1351

RESULT 2  
ABS76635 standard; DNA; 2197 BP.  
ID ABS76635;  
AC ABS76635;  
DT 11-DEC-2002 (first entry)  
XX  
DE DNA encoding novel human metalloproteinase MPI.  
XX  
XX Metalloproteinase; MP-1; immune disorder; glutamate transport; cancer;  
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
KW neurological disorder; gene; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200272751-A2.  
XX  
XX 19-SEP-2002.  
XX  
XX 05-FEB-2002; 2002WO-US003353.  
XX  
XX 05-FEB-2001; 2001US-0266518P.  
PR

PR 10-APR-2001; 2001US-0282814P.  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S;  
XX WPI; 2002-723329/78.  
XX P-PSDB; ABG96478.  
XX  
XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
XX treating, or ameliorating diseases associated with aberrant  
XX metalloproteinase activity, e.g. immune, metabolic, inflammatory and  
XX neurological disorders.  
XX  
XX Claim 1; Fig 1A-C; 473bp; English.  
XX  
XX The invention describes an isolated nucleic acid molecule (I) encoding a  
XX metalloproteinase (MP-1). (I) is useful for preventing, treating, or  
XX ameliorating a medical condition, particularly an immune disorder, an  
XX aberrant glutamate transport or motor neuron disorder, such as  
XX amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
XX condition. The compositions and methods are also useful for diagnosing,  
XX prognosticating, treating, ameliorating and/or treating disorders  
XX associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
XX disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
XX aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
XX or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
XX Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
XX and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
XX or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
XX fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
XX Alzheimer's disease or Parkinson's disease). This sequence represents a  
XX metalloproteinase MP1 polynucleotide  
XX  
XX  
XX Sequence 2197 BP; 681 A; 441 C; 439 G; 636 T; 0 U; 0 Other;  
XX  
XX  
XX Alignment Scores:  
XX Pred. No.: 2,146-212 Length: 2197  
XX Score: 2125.00 Matches: 414  
XX Percent Similarity: 100.00% Conservative: 0  
XX Best Local Similarity: 100.00% Mismatches: 0  
XX Query Match: 100.00% Indels: 0  
XX DB: Gaps: 0  
XX  
XX US-10-649-273-2 (1-414) x ABS76635 (1-2197)  
XX  
XX QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheIysProSerLysArgLysValTyr 20  
XX DB 231 ATGCTAAATCTTGACATAGCTGCAGAGATTTTAAAAACCATCAAAAGAAAGTTTAT 290  
XX QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysValLeu 40  
XX DB 291 GAATTTTAAAGAGATTTTAAATTTTCACTCGAACAATATTTCTTCAATAAATAGATATG 350  
XX QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaValAspGluThrGlyAsnVal 60  
XX DB 351 GGAATTTGAATAGTGTGATGATATACAGACGCTGCTGTGTGATGAATCACTGAAATGTG 410  
XX QY 61 LeuGlyIuAlaIleHisSerGlnThrGluValHisLysLeuValGlyIleValPro 80  
XX DB 411 TTGGAGAGAGCAATATCTCCAACTGAACTTATTTAAAAACAGGTGGATTTCTTCT 470  
XX QY 81 ProHlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
XX DB 471 CCAGACGCTCAACACCTTCAAGAGAAATATTTCAACGAAATGTACAGAACCTCTTCT 530  
XX QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
XX DB 531 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATTTGCACTTAAACCAAGACACTT 590  
XX QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140  
XX DB 591 GCTTTAAGCCTGGAGGTGGGCTTATCATTTAGCTTACAGCTGTGTAGACAGTTAAAAAG 650

QY 141 ProPheIleProIleHisIsmecGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
DB 651 CCATTCAATCCCATTCATCATATGAGAGCTCATGCACTTACTATTAGGTGACCAATATA 710  
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
DB 711 GTAGAAATTTCTTTTATGATTTCTTTGATTTCTGAGAGTCACTGCTGTGGCATTTAGT 770  
QY 181 GluGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
DB 771 CAAGAGCTTCAAGATTTTCTGCTTCTTGAAAAGCTTTGGACATGACACAGGAGCATG 830  
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220  
DB 831 CTTCAGACAGGTGCAAGAAACATTTCTTTAATAAATCATCAGAGTCTCCACATAGT 890  
QY 221 GlyGlyLysValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
DB 891 GGTGGGAAAGCCATAGAACATTTGGCCCAACAGGAATATGATTTCAATTTGACATCAAA 950  
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB 951 CTTCCTTGCAATCATGCTAAAAATGTGATTTTCTTTTACTGAGACTTCAACAGTTACT 1010  
QY 261 AspLysIleIleMetCysLysGlyLysGlyGlyIleGlyLysGlyGlnIleLeuSer 280  
DB 1011 GATTAATATATATGAAAAAGAAAAAGAAAGAGTATTGAGAAAGGCAAAATCTTCT 1070  
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
DB 1071 TCAGCAGACAGACATTCCTGCCACAGTACACACACATGCAATGCTCATCTTGTGAAGA 1130  
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320  
DB 1131 ACACATCGGCTATTTCTGTTTGTAAAGCAGAGACTGTGTACTCAAAATTAATGCACTA 1190  
QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeu 340  
DB 1191 CTGGTTGCATCTGTGTGTGTCGCAAGTAACTTTATATCCGACAGCTCTCGAAATTTTA 1250  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360  
DB 1251 ACAAAACGCAACACAGTCACTTTGTTGTGTCCTCTCCAGACTATGCACTGATTAATGCG 1310  
QY 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
DB 1311 ATTATGATTTGCAATGGAATGTATTTGAAAGACTAGTGTGGCTGGCATTTTATCATGAC 1370  
QY 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGluVal 400  
DB 1371 ATTAGAAGGCATCCGCTATGAAACCAAAATGTCCTTGGATGATACATCAAAAAGAGTT 1430  
QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
DB 1431 GGAAGAGCTTCCATTAAGTACCAATTTAAAAATGAGATA 1472

RESULT 3  
ABT23207  
ID ABT23207 standard; DNA; 2572 BP.  
XX  
XX AC ABT23207;  
XX  
XX 01-MAY-2003 (first entry)  
XX  
XX Human protein modification + maintenance molecule DNA SEQ ID No 36.  
XX  
XX Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;  
XX cerebroprotective; antiParkinsonian; neurotropic; antiinflammatory;  
XX antitumor; hepatotropic; gynaecological; antibacterial; virucide;  
XX protozoicide; antiparasitic; cell proliferative disease; PMOD;  
XX protein modification and maintenance molecule; immunogenic fragment;  
XX cancer; autoimmune; inflammatory disease; neurological disorder;



Db 1164 AACAAACGACACAGTCACCTTTGTTGTCTCTCCCTCCAGACTATGACATGATTAATGCC 1223  
Qy 361 IIEmerIIeAlATrpaSngIyIlegIuaSgleuArGlaGIyleuGIyIleuHIAsp 380  
Db 1224 ATTATGATTGCATGAAATGGTATGTAAGACATACGAGCTGGCTGGCATTTTACATGAC 1283  
Qy 381 IIEGIGLYIIEArGfYrGluProlYsCYeProlEugIyValAspIIEserLYsGIuVal 400  
Db 1284 ATAGAAGCATCCGCTATGTAACCAAAATGTCCCTTGAGTAGACATATCAAAAGAAGTT 1343  
Qy 401 GIVGluAlSerIIElySValProGIuLeuLYeMeHGuIle 414  
Db 1344 GGAAGAGCTTCATATAAGTACCAATTAATAATGAGATA 1385  
RESULT 4  
AAD46856  
ID AAD46856 standard; cDNA; 1820 BP.  
XX  
AC AAD46856;  
XX  
DT 27-JAN-2003 (first entry)  
XX  
DE Human glycoprotease 28472 cDNA.  
XX  
KW Human; adenosine deaminase; seven transmembrane domain receptor; cancer;  
KW 7TM; glycoprotease; immune disorder; IGA deficiency; allergy; arhythmia;  
KW rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;  
KW hypertension; ischaemic heart disease; obesity; myocardial infarction;  
KW endothelial cell disorder; Grave's disease; psoriasis; brain disorder;  
KW Parkinson's disease; Alzheimer's disease; haematopoietic disorder;  
KW cerebral oedema; metabolic disorder; liver disorder; platelet disorder;  
KW chromosome mapping; tissue typing; gene therapy; neuroprotective;  
KW cytoskeletal; anorectic; cardiac; haemostatic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 146..1390  
FT /tag= a  
FT /product= "Human 28472 protein"  
FT /note= "This region is specifically claimed as SEQ ID NO:  
6 in claim 1 of the specification"  
XX  
PN WO200274960-A2.  
XX  
PD 26-SEP-2002.  
XX  
PF 08-NOV-2001; 2001WO-US051427.  
XX  
PR 08-NOV-2000; 2000US-0246768P.  
XX  
PR 08-NOV-2000; 2000US-0246772P.  
XX  
PR 15-NOV-2000; 2000US-0249185P.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Leidy KR, Kapeller-Libermann R, Glucksmann M;  
DR WPI; 2002-759898/82.  
XX  
DR P-PSDB; AAS29234.  
XX  
PT New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,  
PT useful for diagnosing and treating cancer, immune, cardiovascular,  
PT hemetopoietic, brain, pain, metabolic, liver or platelet disorders, and  
PT in pharmacogenomics.  
XX  
PS Claim 1; Fig 8; 178pp: English.  
XX  
CC The present invention relates to novel 38650, 28472, 5495, 65507, 81588  
CC or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-  
CC protease or seven transmembrane domain (7TM) receptor family members.  
CC Sequences of the invention are useful in diagnosing and treating cancer  
CC or aberrant cellular proliferation and/or differentiation (e.g. colon or  
CC lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid

CC arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,  
CC hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,  
CC myocardial infarction, thrombus) including endothelial cell disorders  
CC (e.g. psoriasis, Grave's disease), haematopoietic disorder, brain  
CC disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),  
CC pain and metabolic disorders (e.g. obesity), liver disorders or platelet  
CC disorders. They are also useful in screening assays, predictive medicine  
CC (e.g. diagnostic assays, prognostic assays, monitoring clinical trials  
CC and pharmacogenetics) and prophylactic and therapeutic methods. The  
CC nucleic acids may also be used in chromosome mapping, tissue typing and  
CC forensic biology and as surrogate markers. Sequences of the invention are  
CC also used in gene therapy. The present sequence is human glycoprotease  
CC 28472 cDNA  
XX  
SQ Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;  
SQ  
Alignment Scores:  
Pred. No.: 1,24e-208 Length: 1820  
Score: 2088.00 Matches: 407  
Percent Similarity: 99.03% Conservative: 3  
Best Local Similarity: 98.31% Mismatches: 4  
Query Match: 98.26% Indels: 0  
DB: 6 Gaps: 0  
US-10-649-273-2 (1-414) x AAD46856 (1-1820)  
Qy 1 MetLeuIIeLeuThyrStrThraIaGIValaPhePheLYeProSerLYsArGlySValTYr 20  
Db 146 ATCTTAATCTTGACTTAAGCTGACGAGAGTTTATTTTAAACATCAAAAAGAAAGTTTAT 205  
Qy 21 GluPheLeuArGSerPheAsnPhenIAsProGIYThrLeuPheLeuHIsLYsIIEValLeu 40  
Db 206 GAATTTTAAAGAGTTTAAATTTTCAATCCGGAACCTAATTTCTCATAAATAGATTTG 265  
Qy 41 GIVYIEGluThrSerCYsAspAspThraIaAlaIaValaYAspGIuThrGIyAsnVal 60  
Db 266 GGAATTGAAACTAGTGTGTGATGATACAGCAGCTGCTGGTGATGAATCGAATAATGTC 325  
Qy 61 LeuGIyGluAlaIIEHsSerGIuThrGIuValaHIsLeuLYeThrGIyIIEValPro 80  
Db 326 TTGGAGAAAGCAATATACATTCCTCAAACTGAAGTCAATTAATAACAGGTGGATGTTCT 385  
Qy 81 ProAlaIAGInGInLeuHIsArGIuAsnIIEGInArGIIEValGInGInAlaLeuSer 100  
Db 386 CCAGCAGCTCAACAGCTTCCACAGAAATATTCACAGAAATATACAGAGCTCTTTCT 445  
Qy 101 AlaSerGIyValSerProSerAspLeuSerAlaIIEAlaThrThrIIElySProGIYLeu 120  
Db 446 GCCAGTGAAGTCTCTCAAGTGACTCTCAGCAATTCGAATTAACCAAGAGACTT 505  
Qy 121 AlaLeuSerLeuGIyValaGIyLeuSerPheSerLeuGInLeuValaGIyGInLeuLYs 140  
Db 506 GCTTTAAGCTGGAGATGGCTTATCATTTAGCTTACAGTGGTAGACAGTTAAAAAG 565  
Qy 141 ProPheIIEProIIEHIsHsMetGIuAlaHIsAlaLeuThrIIEArGLeuThrAsnLYs 160  
Db 566 CCAATCAATCCATATCATATATGAGGCTCATGCACTTAATTAAGTTGAGTCAATTA 625  
Qy 161 ValGIuPheProPheLeuValaLeuLeuIIEserGIyYHIsCYsLeuLeuAlaLeuVal 180  
Db 626 GTAGAAATTCCTTTTATGATTTCTTTGATTTCTGAGGTCACGTCTGTTGGCATAGTT 685  
Qy 181 GInGIyValaSerAspPheLeuLeuLeuGIyLYsSerLeuAspIIEAlaProGIYAspMet 200  
Db 686 CAAGAGATTTCAGATTTTCTGCTTCTTGAAAGTCTTTTGACATACACACAGGTGACAG 745  
Qy 201 LeuAspLYsValaIaArGArGLeuSerLeuIIElySAsProGIYCySerThrMetSer 220  
Db 746 CTTGACAAAGTGCAGAAAGACTTCTTTTAATAAACATTCAGAGTCTCACAGAGT 805  
Qy 221 GIVGIYyAlaIIEGIuHIsLeuAlaLYsGInGIYAsnArGPhenIAsPheAspIIElyS 240  
Db 806 GGTGGAAAGCCATTAACATTTGGCCAAACAGAGAAATAGATTTCAATTTGACATCAAA 865



QY 241 ProProLeuH1SH1eAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
 DB CCTCCCTGGCATCATGCTAAATAATGTATTTTCTTTACTGGACTTCAACCGTTACT 925  
 QY 261 AspLysIleIleMetLysLysGlyLysGlyLysGlyLysGlyLysGlyLysGlyLys 280  
 DB GATAAATAATATGAAAACAGAAACAGAAAGGATGAGAAAGGGCAAAATCCGTCT 985  
 QY 281 SerAlaAlaAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 300  
 DB TCAGCAGCAGACATTCCTGCAACAGTACGACACACATGACATGATCTTGGAAAAGA 1045  
 QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320  
 DB ACACATCGGGCTATCTGTTTGTAAAGCAGAGACTTTACTTAAATAATGACAGTA 1105  
 QY 321 LeuValAlaSerGlyGlyValAlaAspAsnAsnAsnAsnAsnAsnAsnAsnAsn 340  
 DB CTGGTTGATCTGGTGGTGTCCAGTAACTTATATCCGACAGACTCTGGAAATTTTA 1165  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360  
 DB 1166 ACAACGCAACACAGTGCATTGTTGTCTCTCCCAAGCTATGACATGATATATGAC 1225  
 QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 DB 1226 ATTAGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1285  
 QY 381 IleGluGlyIleArgGlyIleProLysCysProLeuGlyValAspIleSerLysGlyVal 400  
 DB 1286 ATGAGAGGATCGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1345  
 QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGlnIle 414  
 DB 1346 GGAGAGAGCTCCATTAAGTACCAACATTAATAATGAGATA 1387  
 DB  
 RESULT 5  
 ACA60887  
 ID ACA60887 standard; cDNA; 1820 BP.  
 XX  
 AC ACA60887;  
 XX  
 DT 08-JUN-2003 (first entry)  
 XX  
 DE Human cDNA 28472 encoding a glycoprotease.  
 XX  
 KW Human; ss; gene; cancer; aberrant cellular proliferation;  
 KW differentiation; immune disorders; heart disorder; brain disorder;  
 KW cardiovascular disorder; endothelial cell disorder; pain disorder;  
 KW hematopoietic disorder; blood vessel disorder; metabolic disorder;  
 KW liver disorder; platelet disorder; glycoprotease.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 146..1390  
 FT /tag= a  
 FT /product= "Glycoprotease"  
 FT /note= "This CDS is specifically claimed in claim 1"  
 XX  
 UN US2003009017-A1.  
 XX  
 PD 09-JAN-2003.  
 XX  
 PF 08-NOV-2001; 2001US-00012140.  
 XX  
 PR 08-NOV-2000; 2000US-0246768P.  
 PR 08-NOV-2000; 2000US-0246772P.  
 PR 15-NOV-2000; 2000US-0249185P.  
 XX  
 PA (LEIBY/) LEIBY K R.  
 PA (KAPLE/) KAPLEBER-LIBERMANN R.

PA (GLUC/) GLUCKSMANN M A.  
 XX  
 PI Leiby KR, Kapeller-Libermann R, Glucksmann MA;  
 XX  
 DR WPI: 2003-428888/40.  
 DR P-PSDB; AB009569.  
 XX  
 PT New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid  
 PT molecules, useful for diagnosing, treating cancer, pain, or immune,  
 PT heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic  
 PT and liver disorders.  
 XX  
 PS Claim 2; Fig 8; 90pp; English.  
 XX  
 CC The invention relates to an isolated 38650 (encoding adenosine  
 CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7  
 CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or  
 CC a sequence which is at least 60% identical to the six nucleic acids or  
 CC their open reading frames, fragments of at least 15 nucleotides,  
 CC naturally occurring variants, or a DNA insert of the plasmid deposited  
 CC with the American Type Culture Collection as Accession No. not defined in  
 CC the specification, which encodes the amino acid sequence). Also included  
 CC are a host cell containing the nucleic acids (used to produce the  
 CC proteins), the encoded proteins, an antibody that selectively binds to  
 CC the polypeptide, and identifying a compound that binds to/modulates the  
 CC activity of the polypeptide. The nucleic acid molecules, polypeptides and  
 CC methods are useful for diagnosing, treating cancer, aberrant cellular  
 CC proliferation and/or differentiation, immune disorders, heart disorders,  
 CC cardiovascular disorders including endothelial cell disorders, pain  
 CC and metabolic disorders, liver disorders and platelet disorders (many  
 CC examples of these disorders are given in the specification). The present  
 CC sequence is the Human cDNA 28472 encoding a glycoprotease  
 XX  
 SQ Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;  
 XX  
 Alignment Scores:  
 Pred. No.: 1,24e-208 Length: 1820  
 Score: 2088.00 Matches: 407  
 Percent Similarity: 99.03% Conservative: 3  
 Best Local Similarity: 98.31% Mismatches: 4  
 Query Match: 98.26% Indels: 0  
 DB: 8 Gaps: 0  
 US-10-649-273-2 (1-414) x ACA60887 (1-1820)  
 QY 1 MetLeuIleLeuThrIleThrAlaGlyValPhePheLysProSerLysAlaGlyValIle 20  
 DB 146 ATGCTAATCTTGAAGTAAAGTGAAGTGTATTTTAAACATCAAAAGAAAGTTTAT 205  
 QY 21 GlnPheLeuArgSerPheAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 40  
 DB 206 GAATTTTAAAGAGTTTAAATTTTCACTCGAAGCACTAATTTCTCAAAAATGATG 265  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaAlaValAspGluThrGlyAsnVal 60  
 DB 266 GGAATTCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 325  
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80  
 DB 326 TTGGGAGAAACAAATATCCCAAGTCAATTAATAACAGGTGGGATTTGTTCT 385  
 QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 DB 386 CCAGCAGCTCAACAGCTTCACAGAGAAATATTCAGAAATAGTACAAAGAGCTTTTCT 445  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
 DB 446 GCCAGTGGAGTCTCTCAAGTGAATCTTCAGCAATTCGAATCAATCAAAACAGGACTT 505  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140  
 DB 506 GCTTTAAGCTTGGAGTGGGCTTATCATTTAGCTTACAGCTGTTAGACAGTTAAAAAG 565

QY 141 ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIly 160  
Db 566 CCATTCATTTCCCATTCATCATATATGAGGCTCATGCACTTAATGATTGATGACCAATATA 625  
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaVal 180  
Db 626 GTAGAAATTCCTTTTATGATCTTTTGAATTTCTGGAGGTCACGTCTGTTGGCAATTAAT 665  
QY 181 GlngIyValSerAspPheLeuLeuGlyIlySerSerLeuAspIleAlaProGlyAspMet 200  
Db 686 CAAGAGATTTCAGATTTTCTGCTTCTTGAAAGTCTTTGGACATACCAAGGTGACATG 745  
QY 201 LeuAspIlyValAlaArgArgLeuSerIleuIleIyHisIleProGluCysSerThrMetSer 220  
Db 746 CTTCAGAGTGGCAAGAAAGCTTCTTAAATTAACATCCAGAGTGTCCACACATGAGT 805  
QY 221 GlyGlyValAlaIleGluHisIleuAlaIyGlnGlyAsnArgPheHisPheAspIleIy 240  
Db 806 GGTGGAAAGCCATGAAACATTTGGCCAAACAAAGAAATGATTTGATTTGACATCAAA 865  
QY 241 ProProLeuHisHisAlaIlyAsnAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
Db 866 CTTCCCTTGATCATGCTTAATAATGTGATTTTCTTTACTGGACTTCAACACGTTACT 925  
QY 261 AspIlyIleIleMetClyAspGlyIlyGluGluGlyIleGluIlyGlyGlnIleLeuSer 280  
Db 926 GATATAAATATATGAAACAGAAACAGAAAGAAAGTATGAAAGGGGCAATTCCTGTCT 985  
QY 281 SerAlaAlaAspIleAlaAlaIleValGlnHisIleMetAlaCysHisIleValIlyVal 300  
Db 986 TCAGACAGACGATTTGCTGCCACAGTACAGCACACATATGTCATCTTGGAAAGAA 1045  
QY 301 ThrHisArgAlaIleLeuPheCysIyGlnArgAspLeuLeuProGlnAsnAlaVal 320  
Db 1046 AACATCGGGCTATCTGTTTGTAAAGCAGAGAGACTTGTACTTCAAAATATATGACGTA 1105  
QY 321 LeuValAlaSerGlyGlyValAlaIleSerAsnPheIyIleArgArgAlaLeuGluIleLeu 340  
Db 1106 CTGCTGATCTGCGTGCTGCCAAGTAACTTCTATATCCGACAGCTTCGAAATTTTAA 1165  
QY 341 ThrAsnAlaIleGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360  
Db 1166 ACAAAACGACACAGATGCACTTGTGTGTCTCCCTCCACAGCTATGACCTGATTAATGCG 1225  
QY 361 IleMetIleAlaIlePheAsnGlyIleGluArgLeuAlaGlyIleuGlyIleLeuHisAsp 380  
Db 1226 ATTATGATTGCAATGGAATGTATTAAGACATACGCTGGCTTGGCATTTTACATGAC 1285  
QY 381 IleGluGlyIleArgIyGluProIyAspProLeuGlyValAlaAspIleSerIyGlyVal 400  
Db 1286 ATAGAAAGCATCCGCTATGAAACCAAAATGTCCTCTTGAGTACATATCAAAAGAAAT 1345  
QY 401 GlyIyAlaSerIleIyValProGlnLeuIyMetGluIle 414  
Db 1346 GGAGAGCTTCATTAAGTACCAATTAATAATGAGATA 1387  
RESULT 6  
ABSS7020  
ID ABSS7020 standard; cDNA; 1821 BP.  
XX  
AC ABSS7020;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DB cDNA encoding novel human glycoprotease 28472.  
XX  
KW Cancer; aberrant cell proliferation; aberrant cell differentiation;  
KW breast cancer; ovarian cancer; prostate cancer; colon cancer;  
KW lung cancer; immune disorder; heart disorder; cardiovascular disorder;  
KW endothelial disorder; hematopoietic disorder; blood vessel disorder;  
KW brain disorder; pain; metabolic disorder; liver disorder; diabetes;  
KW platelet disorder; carcinoma; sarcoma; leukemia; Hodgkin's disease;

KW autoimmune disorder; hypertension; atherosclerosis; heart failure;  
KW myocardial infarction; ischaemic heart disease; Crohn's disease;  
KW Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;  
KW cerebral ischaemia; peripheral neuropathy; Alzheimer's disease;  
KW Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; gene; ss.  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 147..1391  
FT /tag= a  
FT /product= "Glycoprotease 28472"  
FT /note= "Specifically claimed in claim 1"  
XX  
PD WO200277233-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 08-NOV-2001; 2001WO-US046724.  
XX  
XX 08-NOV-2000; 2000US-0246768P.  
PR 08-NOV-2000; 2000US-0246772P.  
PR 15-NOV-2000; 2000US-0249185P.  
PA (MILL-) MILLENNIUM PHARM INC.  
PI  
PI Leiby KR, Kapeller-Libermann R, Gluckmann M;  
DR WPI; 2003-029938/02.  
DR P-PSDB; ABG71162.  
XX  
XX  
PT New adenosine deaminase, glycoprotease and seven transmembrane domain  
PT nucleic acids and polypeptides, designated 3650, 28472, 5495, 65507,  
PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or  
PT hypertension.  
XX  
XX Claim 2; Fig 8A-B; 178pp; English.  
XX  
CC The invention describes isolated 3650, 28472, 5495, 65507, 81588 and  
CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The  
CC 3650 nucleic acid molecule comprises a sequence encoding adenosine  
CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding  
CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise  
CC sequences that encode a human seven transmembrane domain (7TM). The  
CC 3650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide  
CC sequences are useful for diagnosing, preventing or treating a subject  
CC with or at risk of developing a disorder, e.g. cancer or aberrant  
CC cellular proliferation and/or differentiation (e.g. breast, ovarian,  
CC prostate, colon or lung cancer), immune disorders, heart disorders,  
CC cardiovascular disorders, endothelial disorders, hematopoietic disorders,  
CC blood vessel disorders, brain disorders, pain and metabolic disorders,  
CC liver disorders or platelet disorders. These disorders include carcinoma,  
CC sarcoma, leukemia, Hodgkin's disease, autoimmune disorders,  
CC hypertension, atherosclerosis, heart failure, myocardial infarction,  
CC ischaemic heart disease, Crohn's disease, Grave's disease, Kawasaki  
CC syndrome, Raynaud's disease, aneurysm, cerebral ischaemia, peripheral  
CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,  
CC cachexia or diabetes. This sequence encodes the novel human glycoprotease  
CC 28472  
XX  
SQ Sequence 1821 BP; 543 A; 365 C; 394 G; 518 T; 0 U; 1 Other;  
XX  
Alignment Scores:  
Pred. No.: 1,24e-208 Length: 1821  
Score: 2088.00 Matches: 407  
Percent Similarity: 99.03% Conservative: 3  
Best Local Similarity: 98.31% Mismatches: 4  
Query Match: 98.26% Indels: 0  
DB: 10 Gaps: 0  
US-10-649-273-2 (1-414) x ABSS7020 (1-1821)  
QY 1 MetLeuIleLeuThrIleArgIyValAlaPhePheIyProSerIyAspGlyValIy 20



QY 21 GIUphelLeuArgSer-PheAsnPhenHisProGlyThrLeuPheLeuHisIleValLeu 40  
 Db 404 GAATTTTAAAGAGTTTAAATTTTCACTCGAAGCACTATTTCTTATATAAAATGATG 463  
 QY 41 GIYIleGluThrSerCysAspSerThrAlaAlaValValAspGluThrGlyAsnVal 60  
 Db 464 GGAAATTGAATCTAGTTGTGTATGATACAGACGCTGCTGTGTGATGAATGAATCTGAAATGTG 523  
 QY 61 LeuGIYgluAlaIleHisSerGlnThrGluValHisIleLeuIleThrGlyIleValIPro 80  
 Db 524 TTGGAGAAAGCAATACATTCCTCAAACTGAAGTTCAATTTAAAAACAGGTGGATTTGTTCTCT 563  
 QY 81 ProAlaIleGlnGlnLeuHisIleArgGluAsnIleGlnArgIleValIleGlnGlnAlaLeuSer 100  
 Db 584 CCAGACGCTCAACAGCTTCACAGAGAAATATTCACGAAATAGTACAGAAAGCTCTTTCT 643  
 QY 101 AlaserGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIleYsProGlyLeu 120  
 Db 644 GCCAGTGGAGTCTCTCCAAAGTACCTCTCAGCAATTCACATTAACATTAACAGAGACTT 703  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuIleYsIlys 140  
 Db 704 GCTTTAAGCTCGAGTGGGCTTATCATTTAGCTTACAGCTGTAGACAGCTTAAATAAG 763  
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIlys 160  
 Db 764 CCATTCATTTCCATTCATCATATATGAGGCTCATGCACTACTATAGTTGATGACCAATAA 823  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyIleHisCysLeuLeuAlaLeuVal 180  
 Db 824 GAGAAATTTCTTTTATGTTCTTTGATTTCTGGAGGTCACGTCGTGTGGATTAAGTT 883  
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyIleYsSerLeuAspIleAlaProGlyAspMet 200  
 Db 884 CAAAGAGTTTCAGATTTTCTGCTCTTGGAAAGCTCTTGGACATAGACCAAGGTGACATG 943  
 QY 201 LeuAspIleValAlaArgArgLeuSerLeuIleYsHisProGluCysSerThrMetSer 220  
 Db 944 CTTCACAGAGTGGCAAGAGACTTCTTTAATTAACATCCAGAGTCTCCACATGAGT 1003  
 QY 221 GlyGlyIleValAlaIleGluHisIleLeuAlaIleGlnGlyAsnArgPheHisPheAspIleYs 240  
 Db 1004 GGTGGAAAGCATATGAAATTTGGCCAAACAGAAATATGATTTGATTCATTCACATCAAA 1063  
 QY 241 ProProLeuHisHisAlaIleYsAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
 Db 1064 CCTCCCTTCATCATAGCTAAATATGATTTCTTTTCTTACCTGAGCTTCAACAGCTTACT 1123  
 QY 261 AspIleValIleMetIleYsGlnIleGluGlyIleGluIleYsGlyGlnIleLeuSer 280  
 Db 1124 GATTAATATATATGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1183  
 QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisIleHisMetAlaCysHisIleValIleYsArg 300  
 Db 1184 TCAGCAGACAGATTCCTCCACAGTACAGCACACATAGGCATGTCATCTTGGAAAGAA 1243  
 QY 301 ThrHisArgAlaIleLeuPheCysIleYsGlnArgAspLeuLeuProGlnAsnAlaValI 320  
 Db 1244 ACAACATCGGGCTATTTCTGTTTGTAGAGCAGAGAGACTTGTATCTCAAAATATATGACAGA 1303  
 QY 321 LeuValAlaSerGlyValAlaIleSerAsnPheYrIleArgArgAlaLeuGluIleLeu 340  
 Db 1304 CTGCTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1363  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360  
 Db 1364 ACAACGCAACACAGGACTTGTGTGTCTCTCCCTCCACAGCTAGTACCTGATATGAG 1423  
 QY 361 IleMetIleAlaIleArgIleGluIleGluIleGluIleGluIleGluIleGluIleGluIle 380  
 Db 1424 ATTAAGATTCGA----- 1435  
 QY 381 IleGluGlyIleArgIleGluIleProIleYsCysProLeuGlyValAspIleSerIleGluVal 400

Db 1436 -----TGATGTCCTCTTGAAGTACATATCAAAAGAGTT 1471  
 QY 401 GIYIleGluAlaSerIleYsValProGlnLeuIleYsMetGluIle 414  
 Db 1472 GAGAGACTTCCATTAATAGTACCAATTTAAATATGAGATTA 1513  
 RESULT 8  
 ADQ24627  
 ID ADQ24627 standard; DNA; 2890 BP.  
 XX  
 AC ADQ24627:  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7447.  
 KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
 KW db.  
 OS Homo sapiens.  
 XX  
 PN W02004048938-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 26-NOV-2003; 2003MO-US038193.  
 XX  
 PR 26-NOV-2002; 2002US-0429739P.  
 XX  
 PA (PROT-) PROTEIN DESIGN LABS INC.  
 XX  
 PI Aziz N, Ginsburg WM, Zlotnik A;  
 XX  
 DR WPI; 2004-441208/41.  
 XX  
 PT Early detection of soft tissue sarcoma comprises determining expression  
 of a gene in a first soft tissue sample and a normal soft tissue sample  
 and comparing the gene expression, also useful in treating soft tissue  
 sarcoma.  
 PT  
 PS Example 2; SEQ ID NO 7447; 210bp; English.  
 XX  
 CC The invention relates to a novel method for detecting soft tissue sarcoma  
 CC which comprises obtaining a first soft tissue sample from an individual  
 CC and a normal soft tissue sample from the same or different individual,  
 CC determining the expression of a gene in both samples and comparing the  
 CC expression of the gene in both soft tissue samples, where a higher level  
 CC of protein expression in the first soft tissue sample indicates the  
 CC presence of soft tissue sarcoma. The method of the invention has  
 CC cyrostatic applications and may be useful for detecting soft tissue  
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
 CC acid sequences may be useful in diagnostic and screening applications.  
 CC The current sequence is that of a human soft tissue sarcoma-upregulated  
 CC DNA of the invention. The current sequence is not shown within the  
 CC specification per se but was submitted in CD format by the inventor.  
 XX  
 SQ Sequence 2890 BP; 869 A; 609 C; 611 G; 789 T; 0 U; 12 Other;  
 Alignment Scores:  
 Pred. No.: 3,2e-193 Length: 2890  
 Score: 1944.00 Matches: 386  
 Percent Similarity: 93.24% Conservative: 0  
 Best Local Similarity: 93.24% Mismatches: 4  
 Query Match: 91.48% Indels: 24  
 DB: 12 Gaps: 1  
 US-10-649-273-2 (1-414) x ADQ24627 (1-2890)  
 QY 1 MetLeuIleLeuThrIleYsThrAlaGlyValAlaPhePheIleYsArgIleYsValIleYr 20  
 Db 1001 ATGCTAATCTTGACTAAGATCGCAGAGGTTTCTTTTAAACATCAACAAAGAAAGTTAT 1060

```

Qy      21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40
Db      1061 GAATTTTAAAGAGTTTAAATTTTCATCTCGAACACATATTTCTCATTAATAATGATATTG 1120
Qy      41 GlyIleGlyThrSerCysAspAspThrAlaAlaIleValIleAspGlyThrGlyAsnVal 60
Db      1121 GGAATTGAACCTAGTGTGATGATACAGACAGCTGCTGTGTGATGAACTGGAAATATGAG 1180
Qy      61 LeuGlyGlyAlaIleHisSerGlnThrGlyValHisIleLeuPheThrGlyGlyIleValPro 80
Db      1181 TTGGAGAGAGCAATACATTTCCAAACCTGAAGTTCATTATAAACAAGTGGAGATTGTTCT 1240
Qy      81 ProAlaIleGlnGlnLeuHisArgGlyAsnIleGlnArgIleValGlnGlnIleAsnSer 100
Db      1241 CCAGAGCTCAACAGCTTCCACAGAGAAATATTCACGAAATGATACAAAGACCTCTTTCT 1300
Qy      101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLeu 120
Db      1301 GCCAGTGAAGTCTCTCAAGTGAACCTCTCAGCAATTGCAACCTAATAAACACAGGACTT 1360
Qy      121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuIleVal 140
Db      1361 GCTTTAACCTGGAGTGGCTTATCATTTACCTTACAGCTGATGAGACAGTTAAAAAG 1420
Qy      141 ProPheIleProIleHisIleMetGlyAlaHisAlaLeuThrIleArgLeuThrAsnIle 160
Db      1421 CCATTCAATCCCATTCATCATATGAGAGCTCATGCACTTACATTTAGCTTGACCAATTA 1480
Qy      161 ValGluPheProPheLeuValLeuLeuIleSerGlyIleHisCysLeuLeuAlaLeuVal 180
Db      1481 GTAGAAATTTCTTTTAACTTTTGAATTTCTTGAATTTCTGAGAGTCACTGCTGTGGCATTA 1540
Qy      181 GlnGlyValSerAspPheLeuLeuGlyIleValSerLeuAspIleAlaProGlyAspMet 200
Db      1541 CAAGAGATTTCAGATTTTCTGCTTCTTGAAAGTCTTTTGACATACACACAGGTGACATG 1600
Qy      201 LeuAspIleValAlaArgArgLeuSerLeuIleIleHisProGlyCysSerThrMetSer 220
Db      1601 CTGACAGAGTGGCAAGAGACTTCTCTTAATAAATCCAGAGTCTCCACCATAGAT 1660
Qy      221 GlyGlyIleValAlaIleGlyHisIleValAlaIleValGlnGlyAsnArgPheHisPheAspIle 240
Db      1661 GGTGGAAAGCCATAGAACATTTGGCCAAACAAAGAAATAGATTTCAATTTGACATCAAA 1720
Qy      241 ProProLeuHisHisAlaIleValAsnCysAspPheSerPheThrGlyLeuGlnHisIleVal 260
Db      1721 CTTCCCTTGATCATGCTGCTAAAAATTTGATTTTCTTTTAACTGACCTTCAACACGTTACT 1780
Qy      261 AspIleValIleIleMetIleValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 280
Db      1781 GATTAATAATATATGAAAAAGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1840
Qy      281 SerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisIleValIleVal 300
Db      1841 TCAGCAGCAGACATCTCTCCACAGACACACACACACATGCACTGCTGTGAAAAAGA 1900
Qy      301 ThrHisArgAlaIleLeuPheCysGlyGlnArgAspLeuLeuProGlnAsnAlaAlaIle 320
Db      1901 ACACATCGGGCTATTTCTGTTTGTAAAGACAGACACTTGTACCTCAAAATATATGACGTA 1960
Qy      321 LeuValAlaSerGlyGlyValAlaIleSerAsnPheThrIleArgArgAlaLeuGlyIleLeu 340
Db      1961 CTGGTTGACATCTGGTGGTGTCCGCAAGTAATCTTGATCCGACAGCTCTGGAAATTTTA 2020
Qy      341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
Db      2021 ACAAGACGACACAGTGCATTTGTTGTGTCTCTCCACGATTAATGCACTGATTAATGGC 2080
Qy      361 IleMetIleAlaIleThrAsnGlyIleGlyArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db      2081 ATTAATGATTTGCA----- 2092
Qy      381 IleGlyGlyIleArgGlyArgIleProIleCysProLeuGlyValAspIleSerIleValIle 400

```

```

Db      2093 -----NGATGCTCTTGAGTACACATATCAAAAAGAGTT 2128
Qy      401 GlyGlyAlaSerIleIleValProGlnLeuIleValMetGlyIle 414
Db      2129 GAGAAAGCTTCCATTAAGTACCAATTAATAATGAGATTA 2170

RESULT 9
ABX70950
ID ABX70950 standard; cDNA; 1416 BP.
XX
AC ABX70950;
XX
DE 05-MAR-2003 (first entry)
XX
DB Novel human cDNA sequence #175.
XX
KW Human; gene; ss; nervous system disorder; peripheral neuropathy;
KW Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
KW neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
KW autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
KW insulin-dependent diabetes mellitus; anaemia; thrombocytopenia; wound;
KW ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
KW fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
KW coagulation disorder; cancer; tumour; inflammatory disease; septic shock;
KW Crohn's disease; anaphylaxis; proliferation; chemotactic;
KW differentiation; stem cell growth factor; haematopoiesis; chemokinetic;
KW haemostatic; antiinflammatory; expressed sequence tag; EST.
XX
OS Homo sapiens.
XX
PN MO200281731-A2.
XX
PD 17-OCT-2002.
XX
PF 29-JAN-2002; 2002MO-US001222.
XX
PR 30-JAN-2001; 2001US-00774528.
XX
PA (HYSEQ) HYSEQ INC.
XX
PA (GOOD) GOODRICH R W.
XX
PI Tang TY, Liu C, Zhou P, Aeundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;
XX
DR WPI; 2003-058563/05.
XX
PT Novel polypeptide useful for treating neurodegenerative diseases, myeloid
PT or lymphoid cell disorders, bone disorders, mechanical and traumatic
PT disorders, coagulation disorders, and inflammatory diseases.
XX
PS Claim 1; Page; 612pp; English.
XX
CC This invention relates to the cDNA sequences encoding an isolated novel
CC human polypeptide. The protein encoded by the nucleic acid of the
CC invention is useful for treating central and peripheral nervous system
CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic
CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus
CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)
CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopenia)
CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,
CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head
CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;
CC bacterial, viral or fungal infections; allergic conditions such as
CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);
CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's
CC disease, anaphylaxis). The protein may be used to inhibit the growth,
CC infection or function of infectious agents such as bacteria, fungi,
CC viruses, or to effect bodily characteristics, biohythms or circadian
CC cycles of rhythms. The protein may also have
CC proliferation/differentiation, stem cell growth factor, haematopoiesis
CC regulation, immune stimulating or suppressing, chemotactic/chemokinetic,

```

CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory  
 CC activities. The cDNA sequences of the invention are useful for expressing  
 CC recombinant protein for analysis. The present sequence represents a novel  
 CC human cDNA sequence of the invention, this sequence is an expressed  
 CC sequence tag (EST) and was identified using subtractive hybridisation  
 XX  
 SQ Sequence 1416 BP: 441 A; 280 C; 273 G; 422 T; 0 U; 0 Other:

## Alignment Scores:

Pred. No.:	5,31e-173	Length:	1416
Score:	1747.00	Matches:	340
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.71%	Mismatches:	0
Query Match:	82.21%	Indels:	0
DB:	8	Gaps:	0

US-10-649-273-2 (1-414) x ABX70950 (1-1416)

```

QY 74 LyfThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluMetIleGlnArg 93
DB 280 AGAACAGGTGGGATTGTTCTCCACAGCAGCTCAACAGCTTCACAGAAAATATTCAACCA 339
QY 94 ILeValGlnGlnAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113
DB 340 ATAGTACAGAAAGCTCTTCTCCAGTGAAGTCTCTCAAGTGAACCTTCACGAATTGCA 399
QY 114 ThrThrIleLeuProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133
DB 400 ACTACCAATAAACAGAGCTTCTTAAGCTCGAGAGTGGCTTATCATTTAGCTTACAG 459
QY 134 LeuValGlyGlnLeuLeuValSerProPheIleProIleHisIleMetGluAlaHisAlaLeu 153
DB 460 CTGGTAGAGACAGTAAAGCAATTCATTCATTCATCATATGAGAGCTCATGCACTT 519
QY 154 ThrIleArgLeuThrAspLeuValGlnIlePheProPheLeuValLeuLeuIleSerGly 173
DB 520 ACTATTAGGTTGACCAATTAAGTAAAGTAAATTTCTTTTAAATTTCTTGAATTTCTGAGGT 579
QY 174 HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLeuSerLeu 193
DB 580 CACTGTCTGTGGCATTTAGTTCAGAGAGTTTCAGATTTTCTGCTTTGGAAAGCTTTTG 639
QY 194 AspIleAlaProGlyAspMetLeuAspLeuValAlaArgArgLeuSerLeuIleLeuHis 213
DB 640 GACATAGCAGCAGGTCGATGCTTCAAGAGTGGCAAGAAAGCTTTCTTAATTAACAT 699
QY 214 ProGluCysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaGlnGlnIleAsn 233
DB 700 CCAGAGTCTCCACATGATGCTGTGGAAAGCCATAGAACATTTGGCCAAACAGCAAAAT 759
QY 234 ArgPheHisPheAspIleLeuSerProProLeuHisAlaLeuAsnCysAspPheSerPhe 253
DB 760 AGATTTCAATTTGACATCAAACTCTTCATCAAGCTTAATAAATTTGATTTTCTTTT 819
QY 254 ThrGlyLeuGlnHisValThrAspLeuIleIleMetCysLeuGlnGlnGlyIle 273
DB 820 ACTGACCTTCAACAGCTTACTGATTAATAATTAATGAAGAAAGAAAGAGAGATTT 879
QY 274 GluLeuGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisIleThrMet 293
DB 880 GAGAGAGGGCAAAATCTGCTTCACAGCAGACATTTGCTGCACAGTACAGACACAAAG 939
QY 294 AlaCysHisLeuValIleArgThrHisArgAlaIleLeuPheCysLeuGlnArgAspLeu 313
DB 940 GCATGTCATCTTGTAAAGAAACACATCGGCTATTCGTTTGTAAAGAGAGAGACTTG 999
QY 314 LeuProGlnAsnAsnAlaValIleLeuValAlaSerGlyValAlaIleSerAspPheTyrIle 333
DB 1000 TTACTCTCAAAATTAATGACGTACTGATGCTGAGTGGTGGTGCAGAGTAACTTCTATATC 1059
QY 334 ArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProPro 353
DB 1060 CGCAGAGCTCTGAAATTTTAAACAAACCAACACAGTGCATTTGTGTGCTCTCTGCC 1119

```

```

QY 354 ArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAla 373
DB 1120 AGACTAGCAGCTGATTAATGACATTAATGATGATGATGATGATGATGATGATGATGATGAT 1179
QY 374 GlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGluProLeuCysProLeuGly 393
DB 1180 GGCTTGGGCAATTTTACATGACATAGAGGATCGCTATAGAACCAAAATGCTCTTGGGA 1239
QY 394 ValAspIleSerIleGlyValAlaSerIleLeuValProGlnLeuLeuMetGlu 413
DB 1240 GTAGACATATCAAAAGAGTGGAGAGAGCTTCCATTAAGTACCAATTTAAATGAG 1299
QY 414 Ile 414
DB 1300 ATA 1302

RESULT 10
ABQ75508 standard; DNA; 1572 BP.
ID ABQ75508
XX
AC ABQ75508;
XX
DT 07-NOV-2002 (first entry)
XX
DE Murine sialoglycoprotease-like gene sequence SEQ ID NO:7.
XX
KW Murine; mouse; protease; calcium activated neutral protease type 5;
KW CAPN5; tryptase 4; sialoglycoprotease; enzyme; genetic disease;
KW neurological; neuropsychological; psychotic illness; transgenic animal;
KW gene; ds.
XX
OS Mus musculus.
XX
PN M0200245491-A2.
XX
PD 13-JUN-2002.
XX
PF 05-DEC-2001; 2001MO-US046405.
XX
PR 06-DEC-2000; 2000US-0251803P.
PR 06-DEC-2000; 2000US-0251820P.
PR 13-DEC-2000; 2000US-0255971P.
XX
PA (DELTA-) DELTAGEN INC.
XX
PI Allen KD, Leviten MW;
XX
DR MPI; 2002-657389/70.
XX
PT Novel transgenic animal, comprising a disruption in protease target gene,
PT is useful for identifying agents that ameliorates a phenotype associated
PT with a disruption in a protease target gene.
XX
PS Example 3; Fig 7; 62pp; English.
XX
CC The present invention describes a non-human transgenic animal (I)
CC comprising a disruption in a protease target gene (PG) selected from
CC calcium activated neutral protease type 5 (CAPN5) gene, tryptase 4 gene
CC and sialoglycoprotease-like gene. Also described is a targeting construct
CC (II), comprising a first polynucleotide sequence homologous to at least a
CC first portion of PG, a second polynucleotide sequence homologous to at
CC least a second portion of PG and a selectable marker. (II) is useful for
CC producing a transgenic mouse comprising a disruption in a protease target
CC gene, by introducing (II) into a cell, introducing the cell into a
CC blastocyst, implanting the resulting blastocyst into a pseudopregnant
CC mouse, where the pseudopregnant mouse gives birth to a chimeric mouse,
CC and breeding the chimeric mouse to produce the transgenic mouse. (I) is
CC useful for identifying an agent that modulates the expression or function
CC of a protease target gene, by administering an agent to (I) and
CC determining whether the expression or function of the disrupted protease
CC target gene in (I) is modulated. (I) is also useful for testing the
CC efficacy of proposed genetic and pharmacological therapies for human

```



CC genetic diseases, such as neurological, neuropsychological or psychotic  
 CC illness. The present sequence represents murine sialolipoproteinase-like  
 CC gene sequence, which is used in an example from the present invention

XX Sequence 1572 BP; 459 A; 337 C; 340 G; 429 T; 0 U; 7 Other;

# Alignment Scores:

Score:	3.76e-152	Length:	1572
Pred. No.:	1549.00	Matches:	326
Percent Similarity:	84.50%	Conservative:	23
Best Local Similarity:	78.93%	Mismatches:	60
Query Match:	72.89%	Indels:	7
	6	Gaps:	2

US-10-649-273-2 (1-414) x AB075508 (1-1572)

```

QY 1 MetLeuileLeuThrIleThrsAlaGlyValPhePheIysProSerIysArgIysValIyr 20
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 5 ATGCTAATGTTAAGAGAGACAGCAGAGAGCTATCCCAAGCCCCCAAGAGTAAGTTAT 64
QY 21 GluPheLeuArgSerPheAsnPheniSProGlyThrLeuPheLeuHisIysIleValLeu 40
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 65 GCATTTTAAAGAAAGATTAGTGTATCCAGAACTCTCTTGTATTAAGCTGGTCTG 124
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValIleAspGluThrGlyAsnVal 60
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 125 GGAATTTGAAACCGAGTGTATGACACAGAGCGCGTGTGTGATTAACCTGGGAATGTG 184
QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisLeuIleThrGlyIleValPro 80
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 185 CTGGGGAGAGCACTGCATCTCCCAACTCAGTTCACTGTAACAAAGTGGAGTTGCTCT 244
QY 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 245 CCAAGTAGCTCAACAACCTCAGACAGAAATATTCAAGATAGTAAAGAAACTCTTCTCT 304
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrIleIysProGlyIleu 120
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 305 GCCTGTAGATACCCCAAGCCAGATCTCTCAGCAATTCACATCAAAACCGGAGCTG 364
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyIleuIysIys 140
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 365 GCCCTAAGTTGGAGAGTTGGCTTATCTTTCCTTACAGCTAATACAGTTAAABAS 424
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIys 160
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 425 CCATTCATCCGATTCATACATGAGAGCTCAGCAGCTAGCTATAGGCTCACCAATAAA 483
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyIleHisCysIleLeuAlaLeuVal 180
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 484 GTAGATTTNCCCTTTTATGTTCTTTGATTTTGGGGGACACTGCTGTGAKRTTAGTC 542
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyIysSerLeuAspIleAlaProGlyIysPhe 200
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 543 CAAGGTTTTCGATTTCTCTGCTCTTGGAAAGCTCTTTGACATAGACACAGCGAGCATG 602
QY 201 LeuAspIysValAlaIleArgIleuSerLeuIleHisIleProGlyIysSerIleMetSer 220
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 603 CTTCACAAAGTGGCAAGAGACTTTCTTTATCAAAATCTCAGAAATGTTTAAATAGT 662
QY 221 GlyGlyIysAlaIleGluHisIleuAlaIysGlnGlyAsnArgPheHisPheAspIleIys 240
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 663 GGTTGAAAAGCTATAGAAAGTGGCCAAAGACGGAATAGATTTTCAATTTTACATCAT 722
QY 241 ProProLeuHisIleAlaIysAsnCysAspPheSerPheThrGlyLeuGlnHisAlaThr 260
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 723 CCACTTATGAGATCTTAAGATTCGATTTTCTTTCACGGGACTTCAACATTAATCT 782
QY 261 AspIysIleIleMetCysIysGluIysGluGluGlyIleGluIysGlyGlnIleLeuSer 280
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 783 GATTAAGCTTAATTAACACAAAGAAAGAAAGAGGACTTGAAGAGGGCAATCTCTCA 842
QY 281 SerAlaAlaAspIleAlaIleAlaThrValGlnHisIleThrMetAlaCysHisIleuValIysArg 300

```

```

Db 843 TCAGCTGCAGACATTGCTCTCGGTACAGATGCAACAGCGTCCACTTGGCAAGAG 902
QY 301 ThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuLeuProGlnAsnAlaVal 320
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 903 ACACATCGCGCTATCTCTTTTGCAGCAAGAAATTTGCTCTTCCAGCTTAACGAGTA 962
QY 321 LeuValAlaSerGlyIleValAlaSerAsnPhenylleArgArgAlaLeuGlnIleLeu 340
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 963 TTAGTTGTATCTCGAGGTGTGTCAACTTGTATACGGAAGAAACATTTGAAATGTCT 1022
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaArgLeuCysThrAspAsnGly 360
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 1023 GCAATATGCAACGATCAGCGGTGTGTGTCCACT-TCAAGACTGTGCACTGACAAATGGG 1081
QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyIleuGlyIleLeu-HisAs 380
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 1082 CATATGATTGCAATGAAATGGAATGGAATTAAGTCCCTGCTGCTGCTTAAACATGGA 1141
QY 380 P-IleGluGlyIleArgIyrGluProIysCysProLeuGlyVal---AspIleSerIysG 399
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 1142 TGAATAGACATTCGGTTATTAACCAATATCTCTTTGAGTGAAGCATTAATCCGAG 1201
QY 399 IuValGlyIleAlaSerIleIysValProGlnLeu 410
    |||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 1202 AAGTTGGCAGA---AGCTTGCCCATTAATAAAGTTA 1233

RESULT 11
ABS76639
ID ABS76639 standard; DNA; 1526 BP.
XX
AC ABS76639;
XX
DT 11-DEC-2002 (first entry)
XX
DE DNA encoding novel human metalloprotease MP1 fragment #1.
XX
KW Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
KW motor neuron disorder; ankyrotrophic lateral sclerosis; AIDS; diabetes;
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
KW encephaloma; cystic fibrosis; vascular disorder; inflammatory disorder;
KW neurological disorder; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200272751-A2.
XX
PD 19-SEP-2002.
XX
PF 05-FEB-2002; 2002WO-US003353.
XX
PR 05-FEB-2001; 2001US-0266518P.
XX
PR 10-APR-2001; 2001US-0282814P.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
CH Chen J, Feder J, Nelson TC, Duclos F, Kryetsek S,
XX
DR P-PSDB; ABG36487.
XX
WP1; 2002-723329/78.
XX
PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,
PT treating, or ameliorating diseases associated with aberrant
PT metalloprotease activity, e.g. immune, metabolic, inflammatory and
XX neurological disorders.
XX
PS Disclosure; Page 462-463; 473pp; English.

```



CC The invention describes an isolated nucleic acid molecule (I) encoding a metalloprotease (MP-1). (I) is useful for preventing, treating, or ameliorating a medical condition, particularly an immune disorder, an aberrant glutamate transport or motor neuron disorder, such as amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like condition. The compositions and methods are also useful for diagnosing, prognosticating, treating, ameliorating and/or treating disorders associated with MP-1 activity, e.g. diabetes, cancer, reproductive disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome, or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease, Parkinson's disease, Huntington's disease or Tourette syndrome), liver and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic fibrosis), and vascular, inflammatory and neurological disorders (e.g. Alzheimer's disease or Parkinson's disease). This sequence represents a metalloprotease MP1 polynucleotide

XX Sequence 1526 BP; 484 A; 297 C; 289 G; 456 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5_9e-135	Length:	1526
Score:	1385.00	Matches:	267
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	65.18%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2 (1-414) x ABS6639 (1-1526)

QY 148 MetGUAIAHIAIAleuThrlleargleuThrasnlyvalGluGlyValSerAppheleu 167  
 DB 1 ATGGAGGCTCATGACCTTACTTATGAGTTGACCAATAAGAAATTTCTTTTATGTT 60  
 QY 168 LeuLeuIISeSerGlyGlyHISCYSeuLeuAlaleuValGluGlyValSerAppheleu 187  
 DB 61 CTTTGTGATTTCTGGAGGTCTGCTGTTGGCATTAAGTTCAAGAAATTTTCAAGTTTCTG 120  
 QY 188 LeuLeuGlyIYSeSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArg 207  
 DB 121 CTTCTTGGAAAGCTTTTGGACATAGCACAGGTGACATGCTTGACAAAGTGGCAAGAA 180  
 QY 208 LeuSerLeuIIeLYeHISProGlyuCYSeSerThmetSerGlyGlyValAlaIleGluHIS 227  
 DB 181 CTTCTTTAATTAACATCATCAGAGTCTCACCACATGAGTGGGGAAGCAATAGAAACAT 240  
 QY 228 LeuAlaIYSeGInGlyAsnArgPheHISpHeAspIleLYeProProLeuHISHisAlaLYS 247  
 DB 241 TTGGCAAAACAGAAATAGATTTTCATTTTGAATCAAACTCCCTTGATCATGCTTAA 300  
 QY 248 AsnCYAspPheSeSerPheThrGlyLeuGlnHISValThraspIleIleIleMetLYS 267  
 DB 301 AATTGTGATTTTCTTTTACTGCACTTCAACAGCTTACTGATTAATAATTAAGAAAG 360  
 QY 268 GluLYSeGInGlyGlyIleGluLYSeGInIleLeuSeSerAlaAlaAspIleAla 287  
 DB 361 GAAAAAGAGGAAGATTTGAGAAAGGGCAATCTCTCTTCAGACAGCAATGCTGCC 420  
 QY 288 ThrValGlnIISrHmetAlaCYSHISLeuValLYeValYAspIleThraspIleIleLeuPhe 307  
 DB 421 ACAGTACAGCACACATGCGATGATCTTGTGAAAACATCGGCGCTATTCGTT 480  
 QY 308 CysLYSeGInArgAspLeuLeuProGInAsnAspAlaValLeuValAlaSerGlyGlyVal 327  
 DB 481 TGTAAAGAGAGAGCTTTTACTTAATAATGAGTACGTGTGCACTGTGTGTGTC 540  
 QY 328 AlaSerAspPheTYrlleargArgAlaLeuGlnIleLeuThraspIleThrGInCYSeThr 347  
 DB 541 GCAAGTAACTTCTAATTCGCGAGAGCTCTGGAAATTTTAAACAAACCAACAGTGCAC 600  
 QY 348 LeuLeuCYsProProProArgLeuCYSeThraspAsnGlyIleMetIleAlaTrpAsnGly 367  
 DB 601 TTGTTGTCTCCCTCCAGACTATAGCACTGATTAATGCAATTATGATTCATGGAATGCT 660

QY 368 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHISAspIleGluGlyIleArgTYrGlu 387  
 DB 661 ATTGAAGACTACCTGCTGGCTTGGCATTTTACATGACATAGAAAGCATCCGATGAA 720  
 QY 388 ProLYeCYProLeuGlyValAspIleSerIleGluValGlyGlnAlaSerIleLYeVal 407  
 DB 721 CCAAAATGTCCTCTTGGAGTACATATCAAAAGAGTTGAGAGAGCTTCCATTAAGTA 780  
 QY 408 ProGInLeuTYSeMetGluIle 414  
 DB 781 CCACATTAATAATGAGATTA 801

RESULT 12  
 ADE31345/c  
 ID ADE31345 standard; DNA; 3158 BP.  
 XX  
 AC ADE31345;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human diagnostic and therapeutic polynucleotide (dthp), SEQ ID NO 100.  
 XX  
 KW diagnostic and therapeutic polynucleotide; dthp; antiarteriosclerotic;  
 KW antiinflammatory; cerebroprotective; antihypertensive; antidiabetic;  
 KW immunosuppressive; neuroprotective; nootropic; neuroleptic; tranquilizer;  
 KW osteopathic; antiarthritic; antirheumatic; cyostatic; hepatotropic;  
 KW virologic; haemostatic; anti-HIV; antihypertensive; thyromimetic;  
 KW dermatological; antibacterial; fungicide; antiparasitic; anticonvulsant;  
 KW thrombolytic; anticoagulant; anorectic; vasotropic; antileuc;  
 KW gene therapy; protein replacement therapy; human; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN NO2003062376-A2.  
 XX  
 PD 31-JUL-2003.  
 XX  
 PF 13-JAN-2003; 2003WO-US001096.  
 XX  
 PR 16-JAN-2002; 2002US-0349384P.  
 PR 17-JAN-2002; 2002US-0349413P.  
 PR 17-JAN-2002; 2002US-0349946P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JT,  
 PI Yu YJ, Tuason O, Yap PE, Amshey SR, Dam TC, Liu TP, Gerstein EH;  
 PI Peralta CH, Lewis SA, Chen A, Marwaha R, Lan RY, Urashka ME;  
 PI Kristnam SR, Kolluru V, Panesar IS;  
 DR WPI; 2003-636732/60.  
 DR P-PDB; ADE31156.  
 XX  
 PT New human diagnostic and therapeutic polynucleotides and polypeptides,  
 PT useful for diagnosing, treating or preventing e.g. leukemia, brain  
 PT cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke  
 PT or Alzheimer's.  
 XX  
 PS Claim 1; SEQ ID NO 100; 634bp; English.  
 XX  
 CC The invention relates to a novel isolated human diagnostic and  
 CC therapeutic polynucleotide (designated dthp). The novel dthp  
 CC polynucleotide comprises: any of 18 DNA sequences consisting of 195-7798  
 CC base pairs fully defined in the specification; a polynucleotide  
 CC comprising a naturally occurring polynucleotide sequence at least 90%  
 CC identical to the dthp polynucleotide; a polynucleotide complementary to  
 CC the dthp polynucleotide or its polynucleotide which is at least 90%  
 CC identical; or an RNA equivalent of any of the polynucleotides mentioned  
 CC above. The dthp polynucleotides have the following activities:  
 CC antiarteriosclerotic, antiinflammatory, cerebroprotective, antihypertensive,  
 CC antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,  
 CC tranquilizer, osteopathic, antirheumatic, antiparasitic, cyostatic,



Db	1512	GCATTTTACATGACATGAGGCAATCCGCTATGAACCAAAATGTCCTTGGAGTAGACA	1453
Qy	396	leSeLyseIuValgIyIuaIaseIriIeIyValProIInIeubIyemIuIle	414
Db	1452	TATCAAAAGAACTTGGAGAGCTTCATTAANAAGTACCAATTAATAATGAGATTA	1397
RESULT 13			
ABL24633			
ID	ABL24633	standard; DNA; 1601 BP.	
XX	AC	ABL24633;	
XX	DT	26-MAR-2002 (first entry)	
DE	Drosophila melanogaster	genomic polynucleotide SEQ ID NO 25372.	
XX	XX		
KW	Drosophila; developmental biology; cell signalling; insecticide;		
XX	pharmaceutical; gene; ds.		
OS	Drosophila melanogaster.		
PN	MO20011042-A2.		
XX			
PD	27-SEP-2001.		
XX			
PF	23-MAR-2001; 2001MO-US009231.		
XX			
PR	23-MAR-2000; 2000US-0191637P.		
XX			
PR	11-JUL-2000; 2000US-00614150.		
XX			
PA	(PEKE ) PE CORP NY.		
PI	Venter JC, Adams M, Li PWD, Myers EW;		
XX			
XX	WPI; 2001-656860/75.		
DR			
PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
PT	genes from Drosophila and for elucidating cell signaling and cell-cell		
XX	interactions.		
XX			
PS	Claim 1; SEQ ID NO 25372; 21pp + Sequence Listing; English.		
XX			
CC	The invention relates to an isolated nucleic acid detection reagent		
CC	capable of detecting 1000 or more genes from Drosophila. The invention is		
CC	useful in developmental biology and in elucidating cell signalling and		
CC	cell-cell interactions in higher eukaryotes for the development of		
CC	insecticides, therapeutics and pharmaceutical drugs. The invention		
CC	discloses genomic DNA sequences (ABJ16176-ABJ10511), expressed DNA		
CC	sequences (ABH01840-ABH16175) and the encoded proteins (ABB57737-		
CC	ABB2072). The sequence data for this patent did not form part of the		
CC	printed specification, but was obtained in electronic format directly		
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences		
XX			
XX	Sequence 1601 BP; 344 A; 477 C; 441 G; 339 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. NO.:	1.51e-64	Length: 1601	
Score:	714.50	Matches: 156	
Percent Similarity:	58.22%	Conservative: 60	
Best Local Similarity:	42.05%	Mismatches: 140	
Query Matchn:	33.62%	Indels: 15	
DB:	4	Gaps: 6	
US-10-649-273-2 (1-414) x ABL24633 (1-1601)			
Qy	39	ValIeugIyIleGIuThrSerCyAspAspThrAlaAlaValValAspGIuThrGly	58
Db	79	GTCTCGGCGCATCGAGACCTCTCGGACACACCGGCAATCCATCGTGACACACCGGC	138
Qy	55	AsvAlIeugIyIyIuaIaIeIhIeSerGIuThrIuValIuIseuIyIthrGIyIle	78
Db	139	CGAGTGATGCGCAATGTCGAGTCCGACACAGAGATTCCACACCCGCTATGAGGCAAT	198

QY	79	ValProAlaIaIaGlnIleuHisArgGluAsnIleGlnArgIleValGlnGluA	98
DB	199	ATTCGGCCCAAGGAGCCAGACCTTCAACGGCCCGCATGAGTCCGGCTCAACAGGCGTGC	258
QY	99	LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuPro	118
DB	259	ATGAGAGCGCGCCCAATTGAAAGCCGAGCAACCACTGACGGCCATGCGGTGACCAACGGTCCC	318
QY	119	GlyLeuAlaLeuSerIleuGlyValGlyIleuSerPheSerLeuGlnIleuValGlyIleu	138
DB	319	GGAATCCCGCTAGATTGCTGTGTGGCGGTGGCCCTTGGACGGCACTGGCCCGCTGGCTGC	378
QY	139	LyAspProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThr	158
DB	379	CAGAAGCCCTGCTGCGCTTACCAACATGAGAGGACACGGCTGAGAGGCCCGCATGGA	438
QY	159	Asn-----LysValGluPheProPheLeuValLeuLeuIleSerGlyHisCysLeu	176
DB	439	CATCCGAGACAGATAGGCTAACCCCTTCTCTCTGCTGGCCAGGGGGGCACTGTGAG	498
QY	177	LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla	196
DB	499	TTGGTGTGTGGCTTAAGGCCCCCGGTGGCTTACGCTTTTGACCAACCTGATATGCG	558
QY	197	ProGlyAspMetLeuAspLysValAlaArgArgLeuSerIleIleHisProGluCys	216
DB	559	CCCGGAGAGGCTTTGACAAAGATCGGTGGCGACCTTACGCTTGACATTTCTGCGGAGTAC	618
QY	217	SerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLys--GlnGlyAsnArgPhe	235
DB	619	CGCTTGTGGAAAGAGAGAGAGCGGCCATTGACATGCGCGCCAGCTGGCAGAGATCCGCTG	678
QY	236	HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly	255
DB	679	GGCTACAGATTCCCGCTGCACTTGGCCCAAGCAGCGGAACTGACATTAGCTTGCCGGCG	738
QY	256	LeuGlnHisValThrAspLysIleIleMetLysGluLysGluGluGlyIleGluLys	275
DB	729	ATCAAGAACACTCTTCGCGGCGCATCGAGAGCGGAGCGGAGCGGAGCGGACACTCGG	798
QY	276	GlyGlnIleLeuSerSerAlaIleAspIleAlaIleAlaThrValGlnHisThrMetAlaCys	295
DB	799	GACGAGATCATCAGCACTACGACGCGACATTCGCGCGGCGCTGCGCTGCTGACGCGG	858
QY	296	HisLeuValLysArgThrHisArgAlaIleLeuPheCys-----LysGlnArgAspLeu	313
DB	859	CACCTGATGACACCGACGAGCGGGGCAATTGATGATCTCTCTGCGCCACAGAGAGCTC	918
QY	314	LeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIle	333
DB	919	TTCCGTGACACCCCGCCACCTGTGTCATGTCCGGGTGTGGCAACAGATGCCATA	978
QY	334	ArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPro	353
DB	979	TACGCCAACATCGAATCTGCGCGCGGAGTATGCGCTGCAGAGCTTCCATCGAAG	1038
QY	354	ArgLeuCysThrAspAsnGlyIleMetIleAlaTyrAsnGlyIleGluThrLeuAla	373
DB	1039	CGGTACTCTCGACACGGGCTGATGATCGCTGGACGGCGTTGAGCAACTG-----	1092
QY	374	GlyLeuGlyIleLeuHisAspIleGlu--GlyIleArgTyrGluProLysCysProLeu	392
DB	1093	-----CTGCAGATTAAGAGGCGACAGCGGCTACGACTAGAC-----	1131
QY	393	GlyValAspIleSerLysGluValGlyGluA	403
DB	1132	AGCATGTATATCAGGCGACGCGGAGTTGCC	1164

RESULT 14  
 AAH15110/C  
 ID AAH15110 standard; cDNA; 1385 BP.  
 AC AAH15110;



QY 204 ----- 204  
Db 207 TTTCAATTATTCCTTGCATCTTTTCGTTTCACAGATATTATTATTAAGACTTAAAA 148  
QY 205 -----AlaArgArgLeuSerLeuIleLeuHisProGluCysSer 217  
Db 147 AATATGTTCTTGTAGTAGGTGGCAGAGAGACCTTTCTTAATATAACATCCAGAGTGTCTCC 88  
QY 218 ThrMetSerGlyGlyValAlaIleGluHisLeuValAlaGlnGlyAsnArgPheHisPhe 237  
Db 87 ACCATGAGTGGTGGGAGAAAGCCATAGAACATTTGGCCAAACAGGAATATGATTTCAATTT 28  
QY 238 AspIleLeuProProLeuHisHisAla 246  
Db 27 GACATCAAACTCCCTTGATCATGCT 1  
RESULT 15  
ABL24632/C  
ID ABL24632 standard; DNA; 3656 BP.  
XX ABL24632;  
AC ABL24632;  
XX 26-MAR-2002 (first entry)  
DT 26-MAR-2002 (first entry)  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 25369.  
DE Drosophila; developmental biology; cell signalling; insecticide;  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; de.  
XX Drosophila melanogaster.  
OS Drosophila melanogaster.  
PN WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
PR 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
PR 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
PA (PEKE ) PE CORP NY.  
PI Venter JC, Adams M, Li PMD, Myers EW;  
XX WPI; 2001-656860/75.  
DR WPI; 2001-656860/75.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions.  
XX Claim 1; SEQ ID NO 25369; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
XX ABBS72072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
SQ Sequence 3656 BP; 1035 A; 829 C; 874 G; 918 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 3.5e-61 Length: 3656  
Score: 687.50 Matches: 157  
Percent Similarity: 55.64% Conservative: 60  
Best Local Similarity: 40.26% Mismatches: 139  
Query Match: 32.35% Indels: 34  
DB: 4 Gaps: 7

US-10-649-273-2 (1-414) x ABL24632 (1-3656)  
QY 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGly 58  
Db 2578 GTCTTGCGCATCGAGACTCTCTGCGAGCAGCAGCGGCACTGCGATGGAACACCGGCG 2519  
QY 59 AsnValLeuGlyGluAlaIleHisSerGlnThrGluValHis 72  
Db 2518 CGAGTATTCGCAATGATGCTGAGTCCGCAACAGAGTTCACACCGGTGAGTGTCA 2459  
QY 73 -----LeuysThr-GlyGlyIleValPr 80  
Db 2458 TCTCCGATCGGTCGCCAATGATGAACTGAAATCGAATGCAATGAGGCAATTAATCC 2399  
QY 80 oProAlaIleGlnGlnLeuHisAsArgGluAsnIleGlnArgIleValGlnGlnAlaLeuSe 100  
Db 2398 GCCCAGGGCCCGAGACCTTCAACCGCCCGCATGAGTCCGCTTACACCGCTGCATGGA 2339  
QY 100 rAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLe 120  
Db 2338 GCGCGCGCAATTGAAAGCCGGAACCACTGACGGCCATGCGGTGACACACGGCTCCGACT 2279  
QY 120 uAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuGly 140  
Db 2278 GCGCTGAGTTGTGTGGTGGCGCTGCTTGGACGCGACCTGCGCCGCTGCGCAGAA 2219  
QY 140 sProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsn- 159  
Db 2218 GCCCTGCTGCGCGTTACACACATGAGGAGCGACAGCGCTGACGCGCGCATGAAACATCC 2159  
QY 160 ----LyValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAl 178  
Db 2158 GAGCAGATAGCGGTACCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2059  
QY 178 aLeuValGlnGlyValSerAspPheLeuLeuGlyGlySerLeuAspIleAlaProG 198  
Db 2098 GGTGCTTAACGCGCGCGCGCTTACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2039  
QY 198 yAspMetLeuAspLeuValAlaArgArgLeuSerLeuIleLeuHisProGluCysSerTh 218  
Db 2038 CGAGCGCTTTGACAAAGATCGGCGCATCAACGCTTGCATTCGCGGAGTACCGCTT 1979  
QY 218 rMetSerGlyGlyValAlaIleGluHisLeuAlaLeu----GlnGlyAsnArgPheHisP 237  
Db 1978 GTGGAAGAGAGAGCGGCGCATTAAGATCGCGCGCGCGCGCGCGCGCGCGCGCTA 1919  
QY 237 eAspIleLeuProProLeuHisHisAlaLeuAlaLeuAlaLeuAlaLeuAlaLeuAla 257  
Db 1918 CGAGTTCCGCTGCACTTCCCGCAGCAGCGGAACTGCACTTCCGCGCGCATCA 1859  
QY 257 nHisValThrAspLeuIleIleMetLeuValGlyGlnGluGluGlyIleGluGlyGly 277  
Db 1858 GAAACAATCTCTCCGCGCGCATCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1799  
QY 277 nIleLeuSerSerAlaAlaAspIleAlaIleAlaIleAlaIleAlaIleAlaIleAla 297  
Db 1798 AGCATATGAGAACTACGCGCACTTTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 1739  
QY 297 uVallyAspGlnHisArgAlaIleLeuPheCys-----LysGlnArgAspLeuLeuPr 315  
Db 1738 GATGACCGGACCGACCGCGCAATTAAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1679  
QY 315 oGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnAspHeTyrlleArgAr 335  
Db 1678 TGACACCGCGCGCACACTGTCATGTCGCGCGGTGTCGCAACACATGCAATATAGCG 1619  
QY 335 galAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysAspProProAla 355  
Db 1618 CAACATGCAACATTCGCGCGCGAGATAGCTGACAGAGCTTCGCTCATGCAAGAGCTTA 1559  
QY 355 uCysThrAspAsnGlyIleMetIleAlaTPAsnGlyIleGluArgLeuArgAlaGlyLe 375  
Db 1558 CTCTCGACACCGGCTGATGATCGCTGCGACGCGGTGAGCACTG----- 1510

```

Qy      375 uGlyIleuHISAspIleGlu---GlyIleargTYrGluProLysCyProlLeuGlyVa 394
Db      1509 -----CTGCAAGATTAAGAGGCCAGCAGCGCGCTTACGACTACGAC-----AGCAT 1466
Qy      394 1AspIleSerLysGluValGlyGluAla 403
Db      1465 TGATATCCAGGGCAGCGCGGATTCGCC 1438

```

Search completed: November 10, 2005, 17:34:03  
 Job time : 963.3 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: November 10, 2005, 16:53:13 ; Search time 5868 Seconds

(without alignments)  
2685.516 Million cell updates/sec

Title: US-10-649-273-2

Sequence: 1 MLITKTAGVFFKPSKRRVY.....DISKEVGASIKVPLKMEI 414

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp  
-O=/cgm2.1/USPRO\_spool\_p/US10649273/runat\_02112005\_091338\_15562/app.query.fasta\_1.1429  
-DB=EST -QFMT=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bites -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=PCT -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273 @CGN 1 1 8076 @runat 02112005 091338 15562 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSEBLOCK=100 -LONGLOC  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hlc: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_ges1: \*  
9: gb\_ges2: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1827	86.0	2284	3	AK045669 Mus muscu
2	1819	85.6	1622	3	AK011265 Mus muscu
3	1634	76.9	1609	3	BC030671 Mus muscu
4	1447	68.1	870	5	BQ423651 AGENCOURT
5	1353.5	63.7	852	5	BX391919 AGENCOURT
6	1275	60.2	922	5	BQ961028 AGENCOURT
7	1173	55.2	701	2	BE740611 AGENCOURT
8	1158	54.5	749	7	CK982692 AGENCOURT
9	1144	53.8	800	7	CO738006 AGENCOURT

10	1132	53.3	1109	5	BM907988 AGENCOURT
11	1111	52.3	765	5	CF114247 Shu12zomi
12	1097	51.6	640	5	BQ636028 h03d11.Y
13	1045	49.2	730	7	CN823245 Oa.sp1bn
14	1038.5	48.9	661	7	CK833139 AGENCOURT
15	1018.5	47.9	879	5	BQ433135 AGENCOURT
16	1017	47.9	1173	6	CD508917 AGENCOURT
17	997	46.9	723	7	CK942291 AGENCOURT
18	978	46.0	579	7	CK819035 AGENCOURT
19	966	45.5	822	7	CF257246 AGENCOURT
20	963	45.3	577	7	CV028547 AGENCOURT
21	954.5	44.9	808	7	CK792978 AGENCOURT
22	954	44.9	833	5	BUI59877 AGENCOURT
23	951.5	44.8	728	4	BG436856 AGENCOURT
24	935.5	44.0	879	5	BQ256052 AGENCOURT
25	935	44.0	871	2	BF308169 AGENCOURT
26	925	43.5	658	7	CF362328 AGENCOURT
27	923	43.4	811	7	CN232675 AGENCOURT
28	920	43.3	859	5	BQ292595 AGENCOURT
29	916	43.1	831	7	CK455183 AGENCOURT
30	915	43.1	812	5	BQ246489 AGENCOURT
31	912	42.9	878	7	CK456228 AGENCOURT
32	911	42.9	661	7	CF787040 AGENCOURT
33	907	42.7	717	7	CK459683 AGENCOURT
34	906	42.6	597	6	CB272391 AGENCOURT
35	902	42.4	533	4	BM126453 AGENCOURT
36	893	42.0	574	5	BQ130908 AGENCOURT
37	893	42.0	687	7	CF360919 AGENCOURT
38	879	41.4	737	7	CV073257 AGENCOURT
39	867	40.8	616	6	CB438468 AGENCOURT
40	857.5	40.4	789	5	BQ242187 AGENCOURT
41	838	39.4	637	7	CK941819 AGENCOURT
42	831	39.1	853	7	CR444993 AGENCOURT
43	825.5	38.8	806	4	BG333362 AGENCOURT
44	819.5	38.6	792	5	BQ403563 AGENCOURT
45	819	38.5	668	5	BQ285782 AGENCOURT

#### ALIGNMENTS

RESULT 1	AK045669	2284 bp	mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length enriched library, clone:R230219017 product: similar to PUTATIVE STALOGLYCOPROTEASE TYPE 2 (Homo sapiens), full insert sequence.
ACCESSION	AK045669	GI:26337528	
VERSION	AK045669.1	GI:26337528	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subcloning of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,		







```

/putative
ORIGIN

```

Alignment Scores:	
Pred. No.:	2.38e-193
Score:	1819.00
Percent Similarity:	90.34%
Best Local Similarity:	86.78%
Query Match:	85.60%
DB:	3
Length:	1622
Matches:	351
Conservative:	23
Mismatches:	40
Indels:	0
Gaps:	0

US-10-649-273-2 (1-414) x AK011265 (1-1622)

Oy	1	MetLeuLeuThrLysThrAlaGlyValPhePheLysProSerLysValGlyValLys	20
Db	207	ATGCTTAATGTTTAAGAAGGACAGCGAGAGCTATTCCCAAGCCCCCAAGAGCTAAAGTTAT	266
Oy	21	GIuPheLeuLysSerPheAsnPheIleProGlyThrLeuPheLeuHisLysIleValLeu	40
Db	267	GGATTTTAAAGAAATTTAGTGTTCATCCCAAGACTCTCTCTGTGTATTAACGTGTCTG	326
Oy	41	GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal	60
Db	327	GGAAATTGAACCCACAGCTGTGATGACACAGAGCGCTGTGTGTGATGAACCTGGAAATGTG	386
Oy	61	LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyGlyIleValPro	80
Db	387	CTGGGGGAGAGCTGGCACTCCCAAACTCAGGTATCATCGAAACGGGTGGGATTTGTCT	446
Oy	81	ProAlaIleGlnGlnLeuHisIleArgLysAsnIleGlnArgIleValGlnGluAlaLeuSer	100
Db	447	CCAGTAGCTCAACAACCTTCACAGAGAAATTTCAACGAATAGTAAAGAAACTCTTCT	506
Oy	101	AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu	120
Db	507	GCTGTAGATACCCCAAGCGATCTTCAGCAATTCGATCAATCAAAACCGGACGTG	566
Oy	121	AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys	140
Db	567	GCCCTTAACCTTGGGAGTGGCTTATCTCTTGTAGCTTACAGTAGTAAATTCAGTTTAAAAAG	626
Oy	141	ProPheIleProIleHisIleMetGlnAlaHisIleAlaLeuThrIleArgLeuThrAspLys	160
Db	627	CCATTCAATCCGATTCATCACTGAGGCTCAAGCATGACTATTAGGCTCACCAATAA	686
Oy	161	ValGluPheProPheLeuValLeuLeuIleSerGlyGlnHisCysLeuLeuAlaLeuVal	180
Db	687	GTAGAATTTCTTTTGTAGTCTTTTGTATTTCTGGCGGTCACTCGCTGTGGCATTTAGTC	746
Oy	181	GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet	200
Db	747	CAAGGTGTTTCGATTTCTGTGCTCTTGGGAAGCTTTGGCATATAGCACCGAGGCACATG	806
Oy	201	LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer	220
Db	807	CTTGCACAGCTGGCAAGAAAGATTCTTTAATCAAAACATCCAGAAATGTTCTTACATAGCT	866
Oy	221	GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys	240
Db	867	GGTGAAAAAGCTATATGAACAGTTGGCCAAAGACGAAATAGATTTCAATTCATCATAT	926
Oy	241	ProProLeuHisIleAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr	260
Db	927	CCACCTTAGAGAAATGCTAAGAAATTTCCGATTTTCTTTCACGGGACTTCAACATATTA	986
Oy	261	AspLysIleIleMetLysLysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSer	280
Db	987	GATAGCTAATATACACACAGAGAAAAGAGAGCAATTCAGAGAGACCAATCTCTGCA	1046
Oy	281	SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg	300
Db	1047	TCAGGTGCAGACATTTGCTGTGGGTACAGCATGACAGCATGCGTGCACCTTGTCCGAAAGA	1106

Oy		301	ThrhiaargalaleleupePecylsygslarapaleuleuProGlnasanaVal	320
Db		1107	ACACATGCCGTATTCTGTTTGGCACAGCAAAAAAATTGCTCTCCAGCTAACGCAGTA	1166
Oy		321	LenvAlaIaseerglgyValAlaIaseAsnPherylleargagaIaleugulleu	340
Db		1167	TTAGTTCTATCTGGAGGTGTGGCAAGTAACTTGTAATCCGAAGAATTGGAATTGTC	1226
Oy		341	ThrAsnaIathrgIncystrLeuleuCysProProProArgLeuCyethrAspaNgly	360
Db		1227	GCAAAATCAAACGCAGTGCAGCTTGTGTGTGTCACCTCCAAAGACTGTCACATGCAATGCGC	1286
Oy		361	IlewerlleatTPRasnngylliegArqneunrgylaagyleuglyleUeHfEasp	380
Db		1287	ATCATGATGTGATGGAAGAGATTGAAGATTACCTGCTGCTGGCGGCTTTACATCAT	1346
Oy		381	IleglucglylleargtryrgluProLysCysProLeuGlYValaIpIleSerLysegluVal	400
Db		1347	CTAAGAGACATCCGATTTGAACCAAAATGTCTCTTGGAATGACATATCCAGAGAAAGTT	1406
Oy		401	GLYcIuIaaserlleysvalProGlnleulysmetculile	414
Db		1407	GCAGAGCTGCCATMAAAGTACCAGCTTAAAATGCGACTT	1448
RESULT 3				
BC030671				
LOCUS				
DEFINITION			Mus musculus O- $\beta$ -N-galactosyltransferase endopeptidase-like 1, mRNA (CDNA clone IMAGE:1226118), containing frame-shift errors.	
ACCESSION			BC030671	
VERSION			BC030671.1	
KEYWORDS			GI:21040459	
SOURCE			HTC.	
ORGANISM			Mus musculus (house mouse)	
REFERENCE			Bukhariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus	
AUTHORS			Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Kuhlman S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., DiStefano L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stjepanovic M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ueda T.B., Toshiyuki S., Carminci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.C., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Halys S.W., Villallon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield V.S., Krzywinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E., Jones S.J. and Marra M.A.	
TITLE			Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences	
JOURNAL			Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	
MEDLINE			22388257	
PUBMED			12477932	
REFERENCE			2 (bases 1 to 1609)	
AUTHORS			Strausberg R.	
TITLE			Direct Submission	
JOURNAL			Submitted (20-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,	
REMARK			USA NIH-MGC Project URL: http://mgc.nci.nih.gov	
COMMENT			Contact: MGC help desk Email: cgabbs-remail.nih.gov Tissue Procurement: Marcello Bento Soares, Ph.D. CDNA Library Preparation: M. Bento Soares, University of Iowa CDNA Library Arrayed by: The I.M.A.G.E.E. Consortium (LMNL)	

DNA Sequencing by: Institute for Systems Biology  
http://www.systemsbio.org

contact: amadan@systemsbiology.org  
Amy Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Series: IRAX Plate: 66 Row: e Column: 10  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21312463  
This clone has the following problem: frame shifted.

## FEATURES

source

```
1..1609
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1226118"
/tissue_type="Thymus gland, mouse"
/clone_1ib="Soares_thymus_2NMT"
/lab_host="DH10B"
/notes="Vector: pT73-Pac"
```

## ORIGIN

## Alignment Scores:

```
Pred. No.: 1.64e-172 Length: 1609
Score: 1634.00 Matches: 327
Percent Similarity: 83.86% Conservative: 21
Best Local Similarity: 78.80% Mismatches: 40
Query Match: 76.89% Indels: 27
DB: Gaps: 2
```

US-10-649-273-2 (1-414) x BC030671 (1-1609)

```
QY 1 MetLeuIleLeuThlyThAlaGlyAlPhePheLyPProSerlySArglyValIlyr 20
Db 229 ATGCTAAAGTTTAAAGACACAGCAGAGCTATTCACCAAGCCCCCAAGAGTAAAGTTTAT 288
QY 21 GluPheLeuArgSerPheAsnPhelSPROGLYThLeuPheLeuIleValIleu 40
Db 289 GCATTTTAAAGAAATTAAGTTCATCCAGAACTCTCTTGTATTAACGTGCTCG 348
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaValAspGluThrGlyAsnVal 60
Db 349 GGAATTAACACAGCTGTGATGACACAGAGCGCTGTGTGATGATAAATCGGAATGTG 408
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeu-LysThrGlyGlyIleValIle 80
Db 409 CTGGGGGAAGCACTCCCAAACTCAGTTTCAATCTGAAGT-----GGGATTTCTCC 462
QY 80 oProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnAArgIleValGlnGluAlaLeuSe 100
Db 463 TCCAGTACTCAACAACTTCAAGAGAAATTTCAACAAATAGTGAAGAAACTCTTTC 522
QY 100 TAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThIleLysProGlyIle 120
Db 523 TGCCTGTGAATCACCCCAAGGCACTCTCAGCAATTCGACATCAACCGGGACAT 582
QY 120 uAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysIly 140
Db 583 GGCCCTAAGCTTGGAGTTGGCTTATCTTTAGCTTACAGCTAGTAAATCAGTTTAAAA 642
QY 140 sProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIly 160
Db 643 GCCATTCATCCCGATTCATCATGAGGCTCAGCAGCTAGCTATTAGGCTCACCAATA 702
QY 160 sValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db 703 AGTAAATTTCTTTTATTTAGTCTTTTATTTCTGCGGTCTCCTGCTGTGGCATTAAT 762
QY 180 TGlGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspHe 200
```

```
Db 763 CCAAGGTCTTCCGATTTCTCTCTGGAGATCTTTGGACATPAGACCAAGCGACAT 822
QY 200 tLeuAspLyValAlaArgArgLeuSerLeuIleLysHisPProGlyCysSerThrMetSe 220
Db 823 GCTTGACAGGTGGCAAGAAAGACTTTTATTAATCAACATCCAGAAATTTTACATAG 882
QY 220 rGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyValAsnArgPheHisPheAspIleLy 240
Db 883 TGGTGAAAAGCTATAGAACAGTTGGCCAAAGACGAAATGATTCATTTACATCA 942
QY 240 sProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGlyLeuGlnHisValTh 260
Db 943 TCCACCTATGCGAATCTAAGAAATTCGATTTTCTTTCACGAGACTTCAACATATTAC 1002
QY 260 rAspLyIleIleMetLysLySgLyGluGluGlyIleGlyLysGlyGlnIleLeuSe 280
Db 1003 TGATTAAGCTTAATACACACAGAAAGAAAGAAAGAGCATTCAGAAAGGGCAATCTGT 1062
QY 280 rSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysAr 300
Db 1063 ATCAGCTGACAGACATTCCTGCTGGGTACAGATGACAGCAAGCGTGCACCTTGGAAAAG 1122
QY 300 gThrHisArgAlaIleLeuPheCysLySgLyGlnAspLeuLeuProGlnAsnAlaVala 320
Db 1123 AACACATCGCGCTAATCTGTTTTCAGACAGAAATTTGCTCTCCAGCTAACGACAGT 1182
QY 320 tLeuValAlaSerGlyGlyValAlaSerAsnPhetyrIleArgArgAlaLeuGluIle 340
Db 1183 ATTAAGTTGTATCTGAGAGCTTTGCAACTTCTTACATCCGAAAGCATTTGGAATTTGT 1242
QY 340 uThrAsnAlaThrGlnCysThrLeuLeuCySProProArgLeuCySThrAspAsnG 360
Db 1243 CGCAAAATGCAACCGACAGTGCAGTTGTGTCTCCACCTCCAAAGCTGTGCAATGG 1302
QY 360 yIleMetIleAlaIleTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAs 380
Db 1303 CATCATGATTTGCA-----TGATGTCTCTTGGAGTACATATTCAGAGAGT 1315
QY 380 pIleGluGlyIleArgTyrgLupProLySProLeuGlyValAspIleSerLySgLyVa 400
Db 1316 -----TGATGTCTCTTGGAGTACATATTCAGAGAGT 1350
QY 400 tGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
Db 1351 TGCAAGAGCTGCCATTAAGATCCGCAATTAATAATGCACTT 1393
RESULT 4
BQ423651 870 bp mRNA linear EST 23-MAY-2002
LOCUS AGENCOURT_7790948 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6065828
DEFINITION 5' mRNA sequence.
ACCESSION BQ423651
VERSION BQ423651.1 GI:21118966
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 870)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DC/DTP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM13342 row: 1 column: 21
High quality sequence stop: 710.
```







QY 259 |TThAspLysIleIleMeLcLysGluLysGluGluGlyIleGlyLysGlyGlnIleIle 279  
 |||  
 Db 543 TACTGATTAATAATATATGAAAAAGAAAAAGAAAGATTGAGAAAGGCAAAATCTT 602  
 QY 279 |UserSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLys 299  
 |||  
 Db 603 GTCTTCAGCAGCAGCATTTGCTGCACAGTACAGACCAATGGCATGTCATCTGTGAA 662  
 QY 299 |sATGThHisArgAlaAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAsnAl 319  
 |||  
 Db 663 AAGAAA-CATCGGCTATTCTGTGTTTGTAAAGCAGAGAGACTTGTACCTCAAAAATAATGC 721  
 QY 319 |aValLeuValAlaSer-GlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluI 339  
 |||  
 Db 722 AGTACTGGTGCATCTTGGGGGGGTGGCAGTAACCTTCAATACCCGCAAACTTGGAAA 781  
 QY 339 |leuThr-AsnAlaThrGlnCysThr-LeuLeuCysProProArgLeuCysThr-A 358  
 |||  
 Db 782 AATTACAAAACCCAAACGGGGCACCTTTGTTGGGTCCCTCCCAAACTATGACCTGG 841  
 QY 358 |sPaanglyIleMetIleAlaTrp-----AsnGlyIleGluArg--LeuArgAlaGlyL 375  
 |||  
 Db 842 AATAATGG---CATTAATGATGCTGGGGAGGAAGGAATTGAAAAAATAACNTGCTGGCT 898  
 QY 375 |eugly 376  
 |||  
 Db 899 TGGGG 903

RESULT 7  
 LOCUS BE740611 701 bp mRNA linear EST 15-SRP-2000  
 DEFINITION 60159573991 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3949640 5',  
 mRNA sequence.  
 ACCESSION BE740611  
 VERSION BE740611.1 GI:10154603  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NIH-MGC http://mhc.nci.nih.gov/  
 1 (bases 1 to 701)  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: c9abbs-r@mail.nih.gov  
 Tissue Procurement: DCTD/DRP  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov  
 Plate: L1CM814 row: n column: 09  
 High quality sequence stop: 701.  
 Location/Qualifiers  
 1..701  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3949640"  
 /issue\_type="adenocarcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 9"  
 /note="Organ: ovary; Vector: pOT87; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCAAGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3,84e-121 Length: 701  
 Score: 1173.00 Matches: 229  
 Percent Similarity: 99.15% Conservative: 3  
 Best Local Similarity: 97.86% Mismatches: 0  
 Query Match: 55.20% Indels: 2  
 DB: 2 Gaps: 0

US-10-649-273-2 (1-414) X BE740611 (1-701)

QY 125 |GlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysProPheIlePro 144  
 |||  
 Db 2 |GGAGTGGCTTATCATTTAGCTTACCTTACCTGAGTGAAGCAGTTAAAAAGCCATTATCTCC 61  
 QY 145 |IleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPhePro 164  
 |||  
 Db 62 |ATTCAATCAATGAGGCTGACCTTACTTATAGTTGACCAATTAAGATGATTTCTT 121  
 QY 165 |PheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnIleValSer 184  
 |||  
 Db 122 |TTTTAGTTCTTTGATTTCTGAGGTCACTGCTGTGGCATTAAGTCAAGAGTTTCA 181  
 QY 185 |AspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysVal 204  
 |||  
 Db 182 |GATTTCTGCTCTTGGAAAGTCTTGGACATAGCACAGTACACATGCTTGAACAGGTG 241  
 QY 205 |AlaArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAla 224  
 |||  
 Db 242 |GCAGAGAACTTTCTTTAATAAATCCAGAGTGTCCACCATGATGATGGGAAAGCC 301  
 QY 225 |IleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHis 244  
 |||  
 Db 302 |ATGAAACATTTGGCCAAACAGAAATAGATTTCAATTTGACATCAAACTCCCTGCAT 361  
 QY 245 |HisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIle 264  
 |||  
 Db 362 |CATGTAAATAATGTGATTTTCTTTTACCTGACCTTCAACACGTTACTGATTAATAATATA 421  
 QY 265 |MetLys--LysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAla 284  
 |||  
 Db 422 |ATGAAAACAGAAAACAGAGAAAGTATTAAGAAAGGGGCAAAATCTGCTTACAGCAG 481  
 QY 284 |spIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArg 304  
 |||  
 Db 482 |ACATTGCTGCCACAGTACAGACACATGCACTGATCTTGTAAGAAAGAACACATCGGG 541  
 QY 304 |IleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAla 324  
 |||  
 Db 542 |CTATTCTGTTTGTAAACAGAGACTTGTACTCTCAAAAATAATGACATGCTGTGCAT 601  
 QY 324 |erGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAla 344  
 |||  
 Db 602 |CTGTGTGTGTGCGCAAGTAACCTTCAATACCGCAGAGCTCTGGAATTTTAAACAAACGCA 661  
 QY 344 |hrGlnCysThrLeuLeuCysProProArgLeuCys 356  
 |||  
 Db 662 |CACAGTGCATTTGTGTGTCTCTCCACAGACTATGC 699  
 RESULT 8  
 LOCUS CK982692 749 bp mRNA linear EST 16-MAR-2004  
 DEFINITION 4115331 BARC 9BOV Bos taurus cDNA clone 9BOV45\_M02 5', mRNA  
 sequence.  
 ACCESSION CK982692  
 VERSION CK982692  
 KEYWORDS CK982692.1 GI:45500672  
 SOURCE EST.  
 ORGANISM Bos taurus (cow)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovinae; Bos.  
 REFERENCE 1 (bases 1 to 749)  
 Sonstegard, T.S., Van Tassel, C.P., Matukumalli, L.K., Harhay,

ORIGIN



TITLE G.P., Bosak, S., Rubenfield, M. and Gasbarré, L.C.  
Production of EST from cDNA libraries derived from immunologically  
activated bovine gut  
JOURNAL Unpublished (2004)  
COMMENT Contact: Tad S. Sonstegard  
Bovine Functional Genomics Laboratory  
Animal and Natural Resources Institute  
Bldg. 200 Rm2A BARC-East, Beltsville, MD 20705, USA  
Tel: 3015048416  
Fax: 3015048414  
Email: tad@anri.barc.usda.gov

Single pass sequencing. Bases called and trimmed with phred  
0.000925 using options -trim\_alt "-trim.fasta. Vector identified  
by cross\_match using options -mismatch 12 -minscore 18  
Plate: 45 row: M column: 02  
Seq primer: CCCAGTCACGACGCTTCTTAACG  
High quality sequence stop: 749.  
Location/Qualifiers  
1. 749  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/strain="Holstein"  
/db\_xref="taxon:9913"  
/clone="9BOV45\_M02"  
/sex="Male"  
/tissue\_type="Pooled"  
/dev\_stage="Multiple"  
/lab\_host="DH10B T1 phage resistant"  
/clone\_lib="BARC 980V"  
/note="Organ: Abomasum; Vector: pagen-1; Site\_1: EcorV;  
Site\_2: NotI; Equimolar amounts of mRNA extracted from  
fundic and pyloric abomasums of 18 and 21 week old steers.  
Exposure to Osteragis osteragi was initiated at 15 weeks  
of age. fundic and pyloric abomasum"

## ORIGIN

Alignment Scores:  
Pred. No.: 2,09e-119 Length: 749  
Score: 1158.00 Matches: 225  
Percent Similarity: 95.18% Conservative: 12  
Best Local Similarity: 90.36% Mismatches: 10  
Query Match: 54.49% Indels: 2  
DB: 7 Gaps: 0

US-10-649-273-2 (1-414) x CK982692 (1-749)

Qy 49 ThrTAAAlaAlaValaValaAspGluThrGlyAsnValLeuGlyGluAlaIleHisSerGln 68  
Db 1 ACAGCAGCTGCTGTAGTGATGAACTGGAATTTCTTCCAGTACGCTCAACATTCCTCA 60  
Qy 69 ThrGluValHisLeuValThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArg 88  
Db 61 ACTGAGTTCATTTAACACAGGTGGAGTATTTCTCCAGTACGCTCAACATTCCTCA 120  
Qy 89 GluAsnIleGlnArgGlyLeuValGlnGluAlaLeuSerAlaSerGlyValSerProSerArg 108  
Db 121 GAATATATTCACGATAGTACAGAGACTCTCTCTCCAGAGAGACTCTCTCAACTGA 180  
Qy 109 LeuSerAlaIleAlaThrThrIleValProGlyLeuAlaLeuSerLeuGlyValGlyLeu 128  
Db 181 CTCCTCAGAGTTCACACCATTAAGCCAGGACTTGTTCCTTAAGCTTGGCGCTAGGTTTA 240  
Qy 129 SerPheSerLeuGlnLeuValGlyGlnLeuValSerProPheIleProIleHisHisMet 148  
Db 241 TCGTTTACTTACACACTGTGTAGACCAAGTTTAAAGCCCTTCATTCCTCAATTCACAG 300  
Qy 149 GluAlaHisAlaLeuThrIleArgPheThrAsnValGluPheProPheLeuValLeu 168  
Db 301 GAGGCTATGACCTCACTAATAGTTTAAACAAATAGTGAATTCCTGTTTATGTTCTT 360  
Qy 169 LeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 188  
Db 361 TTGATTTCTGAGAGCTATGTCTTTTGGCATTTAGTAGAGGATTTCAGATTTCTTCTT 420

Qy 189 LeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspValAlaArgArgLeu 208  
Db 421 CTGGAGAGTCTTTGGACATAGCACCGAGTGCATGCTTGAACAAGTACGAAGAAGCTT 480  
Qy 209 SerLeuIleValHisPProGlyCysSerThrMetSerGlyGlyValAlaIleGlnHisLeu 228  
Db 481 TCTTATTAATAACATTCACAGGTGTCTCCACATGATGTGGCGGAAGGCTATAGAACATTTG 540  
Qy 229 AlAlaGlnGlyAsnArgPheHisPheAspIleValSerProPheLeuHisAlaValAsn 248  
Db 541 GCCAAACAGGAATATGATTCATTTGATTTCCAGCTCCCAACAGTCTTAATAAT 600  
Qy 249 CysAspPheSerPheThrGlyLeuGlnHisValThrAspValIleMetIleValGlyGlu 268  
Db 601 TGTGATTTCTTTCTTTCTGACCTCAACAGATTATGATTAAGATTAAGTACGAAAGAA 660  
Qy 269 LysGluGluGlyIleGlyGlyGlyGlnIleLeuSerSerAlaAlaAspIleAlaThr 288  
Db 661 AAAGAGAGAGTATCGACAGCGGCGAGTCTCTTCCTTCAGCTGACAGCATTTGCTGCG 720  
Qy 289 ValGlnHisThrMetAlaCysHisLeu 297  
Db 721 GTGCAGCACACCGTGCCTGCACATT 747

RESULT 9  
CO738006 800 bp mRNA linear EST 29-JUL-2004  
LOCUS SLH03C19e02f1 squirrel heart library 1 Sperophilus lateralis cDNA  
DEFINITION  
clone 19e02 5', mRNA sequence.  
ACCESSION  
CO738006  
VERSION  
CO738006.1 GI:50825276  
KEYWORDS  
EST.  
SOURCE  
Sperophilus lateralis (golden-mantled ground squirrel)  
ORGANISM  
Sperophilus lateralis  
Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;  
Sperophilus.  
1 (bases 1 to 800)

REFERENCE  
AUTHORS  
TITLE  
Microarray analysis of transcriptional changes during hibernation  
in the golden mantled ground squirrel, *Sperophilus lateralis*  
Unpublished (2004)  
CONTACT: Andrew R. Cossins  
Laboratory for Environmental Gene Regulation  
University of Liverpool  
School of Biological Sciences, The Biosciences Building, Crown  
Street, Liverpool, United Kingdom, L69 7ZB  
Tel: +44(0)151-795-4510  
Fax: +44(0)151-795-4431  
Email: cossins@liverpool.ac.uk  
Vector has been trimmed from this EST.  
Plate: 19 row: e column: 02  
Seq primer: pfic T7 (5'-AATAGACTCACTAAGG-3')  
High quality sequence stop: 800.  
Location/Qualifiers  
1. 800  
/organism="Sperophilus lateralis"  
/mol\_type="mRNA"  
/db\_xref="taxon:76772"  
/clone="19e02"  
/sex="Male & female"  
/tissue\_type="Heart"  
/dev\_stage="Adult"  
/lab\_host="R.coli Electromax DH10B"  
/note="Vector: pFIC; Site\_1: SalI GTCCAG; Site\_2: BamHI  
GAATC; Normalized and subcloned cDNA library prepared  
from heart of hibernating and summer animals"

## FEATURES

source

## ORIGIN

Alignment Scores:

Pred. No.: 8.74e-118 Length: 800  
 Score: 1144.00 Matches: 226  
 Percent Similarity: 97.05% Conservative: 4  
 Best Local Similarity: 95.36% Mismatches: 7  
 Query Match: 53.84% Indels: 0  
 DB: 7 Gaps: 0

US-10-649-273-2 (1-414) x CO738006 (1-800)

QY 1 MetLeuileuThrlyThrAlaGlyValPhePheLySProSerLySArgLySValTyr 20  
 Db 88 ATGCTAATATTTGAATAAAGACACGCGAGTTTAAAAACATCCAAAAAGAAAGTTGAT 147  
 QY 21 GluPheLeuArgSerPheAsnPhenIleProGlyThrLeuPheLeuHisLySValLeu 40  
 Db 148 GGATTTTAAAGAGTTTAAATTTTATTCCTCGAACAATATTTCTTCAATAATAGATATG 207  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGlyAsnVal 60  
 Db 208 GGTATTTGAACAAAGCTGTGATGATACAGACGCTGGGTGTAGATGAACCTGAAATGTG 267  
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLySThrGlyIleValPro 80  
 Db 268 TTGGAGAAAGCAATACATTCCTCAAACTGAAGTTCAATTAAAAACAGTGGATTATCT 327  
 QY 81 ProAlaIleGlnGluLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 Db 328 CCAGTAGCTCAACAACCTTCATAGAGAAAACATTCACGAATAGTACAAAGAGCTGTTGT 387  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLySProGlyLeu 120  
 Db 388 GCCAGTGGAAATATCTCCAAAGTGAACCTCAGCAGTTGCACTAAAGCCAGAGACT 447  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLyS 140  
 Db 448 GCTTTAAGCTTAAAGAGTAGCTTATCATTTAGTTTCACTGTGTAGAACACAGCGAAATG 507  
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLyS 160  
 Db 508 CCATTCATCCCATTCATCATATGAGAGGCTCAGCACTTACATTAATAGTTGCAATATAA 567  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 Db 568 GTAGAGTTTCCATTTTATGATCTTTGATTTCTGGAGGTCACGTCTCTTGGCATTAAGT 627  
 QY 181 GlnGlyValSerAspPheLeuLeuGlyLySLeuSerLeuAspIleAlaProGlyAspMet 200  
 Db 628 CAAAGGATTTCAGATTTTCTGCTTCTTGAAGCTTTGGACATTTGACACAGGTGACATG 687  
 QY 201 LeuAspLySValAlaIleArgArgLeuSerLeuIleLySHisProGlyCysSerThrMetSer 220  
 Db 688 CTTGACAAAGTACCAAGAGACCTTTCTTAAATCAAAATCCAGATCTCCACCATAGT 747  
 QY 221 GlyGlyLySAlaIleGluHisLeuAlaLySLeuGlyValAsnArgPheHisPhe 237  
 Db 748 GGGGGGAGAGGCTATAGAACATTTGGCCAAACAAAGAAACAGATTTCAATTT 798

RESULT 10  
 BM907988 1109 bp mRNA linear EST 12-MAR-2002  
 LOCUS BM907988  
 DEFINITION AGENCOURT 6707465 NIH\_MGC\_119 Homo sapiens cDNA clone IMAGE:5744934  
 5', mRNA sequence.  
 ACCESSION BM907988  
 VERSION BM907988.1 GI:19358367  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 1109)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
 TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: c9apbs-r@mail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LHM12767 row: b column: 07  
 High quality sequence stop: 602.  
 Location/Qualifiers  
 1..1109

## FEATURES

## source

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5744934"  
 /issue\_type="medulla"  
 /lab\_host="DH108"  
 /clone\_11b="NIH MGC 119"  
 /note="Organ: Brain; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: EcoRV (destroyed); RNA source normal medulla from  
 anonymous male age 27. Library is oligo-dT primed and  
 directionally cloned (EcoRV site is destroyed upon  
 cloning). Average insert size 1.3 kb, insert size range  
 0.9-3 kb. Library is normalized and enriched for  
 full-length clones and was constructed by C. Gruber  
 (Invitrogen). Research Genetics tracking code 013. Note:  
 this is a NIH\_MGC library."

## ORIGIN

## Alignment Scores:

Pred. No.: 3.28e-116 Length: 1109  
 Score: 1132.00 Matches: 241  
 Percent Similarity: 88.89% Conservative: 15  
 Best Local Similarity: 83.68% Mismatches: 23  
 Query Match: 53.27% Indels: 9  
 DB: 5 Gaps: 4

US-10-649-273-2 (1-414) x BM907988 (1-1109)

QY 1 MetLeuileuThrlyThrAlaGlyValPhePheLySProSerLySArgLySValTyr 20  
 Db 129 ATGCTAATCTTGAATAGACTGACAGAGTTTAAAAACATCAAAAGAAAGTTAT 188  
 QY 21 GluPheLeuArgSerPheAsnPhenIleProGlyThrLeuPheLeuHisLySValLeu 40  
 Db 189 GAATTTTAAAGATTTTAAATTTTCATCTCGAACAACATATTTCTTCAATAATAGTATG 248  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGlyAsnVal 60  
 Db 249 GGAATTAACCTAATTTGATGATACAGACGCTGTGTGATGAATGAACCTGAAATGTG 308  
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLySThrGlyIleValPro 80  
 Db 309 TTGGAGAAAGCAATACATTCCTCAAACTGAATTCATTTAAAAACAGGTGGATGTTCT 368  
 QY 81 ProAlaIleGlnGluLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 Db 369 CCAGCAGCTCAACAGCTTCAAGAGAAATATTCACAAAGTAATGACAAAGCTCTTCT 428  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLySProGlyLeu 120  
 Db 429 GCCAGTGAAGTCTCTCCAAAGTACCTTCAGCAATGTGCACTACCATTAACCAAGACTT 488  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLyS 140  
 Db 489 GCTTTAAGCTTGAAGGTGGCTTATCATTTAGCTTACAGCTGTGTAGACAGTTAAAAAG 548  
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLyS 160  
 Db 549 CCATTCATTCATTCATCATATGAGAGGCTCATGCACTTACATTAATAGTTGACCAATATA 608



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
FEATURES  
source

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 640)  
Wistow, G., Bernstein, S.L., Wyatt, M.K., Ray, S., Behal, A., Touchman, J.W., Bouffard, G., Smith, D. and Peterson, K. Expressed sequence tag analysis of human retina for the NEIRBank Project: Retbindin, an abundant, novel retinal cDNA and alternative splicing of other retina-preferred gene transcripts  
Mol. Vis. 8 (4), 196-204 (2002)  
22103461  
12107411  
Contact: Wistow G  
Section on Molecular Structure and Function  
National Eye Institute  
6/331, NIH, Bethesda, MD 20892-2740, USA  
Tel: 301 402 3452  
Fax: 301 496 0078  
Email: graeme@helix.nih.gov  
Plate: 03 row: d column: 11  
Seq primer: M13RP1 reverse primer (ABI).  
Location/Qualifiers  
1. 640  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="hd03d11"  
/tissue\_type="Retina"  
/dev\_stage="Adult"  
/lab\_host="EMDH108"  
/clone\_lib="Human Retina cDNA (Un-normalized, unamplified): hd/he"  
/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue was dissected from two 80 year old donors with no observed eye disease. 100ug of total RNA was used for library construction. A directionally cloned cDNA library in the pSPORT1 vector (Life Technologies) was essentially following Bioerive Biotechnology (Laurel MD) essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetechn.com/). First strand synthesis was carried out using a Not I primer-adaptor  
[5'-GGACTAGTTCTTAGATCGGAGCGGCCCTT)is-3']. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN  
Alignment Scores:  
Pred. No.: 1.22e-112 Length: 640  
Score: 1097.00 Matches: 209  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 51.62% Indels: 0  
DB: 5 Gaps: 0

US-10-649-273-2 (1-414) x BQ636028 (1-640)

QY 206 AAGAGGleSerleuilelyshisProgluCySerThmetSerGlylyAlaile 225  
DB 3 AAGAGGCTTTCTTTAATAAACATCCAGAGTGTCTCCACCATGAGTGTGGAGAAAGCATTA 62  
QY 226 GIuHIsleuAlaLySGInGlyAenArGpHeHIsPhaSpIleLySPoRLeuHIs 245  
DB 63 GAACATTGGCCAAACAGGAATATAGATTTCATTTGACATCAACCTCCCTGCATCAT 122  
QY 246 AAlaYsAsnCysAspHeserPheThrGlyLeuGlnHIsValThrAspLySlelleMet 265  
DB 123 GCTAAATAATGATGATTTCTTTACTGACATCAACCTCTGATTAATAATATATG 182  
QY 266 LysLyGluLyGlnGluGluGlylleGluLySGlyGlnleleuSerSerAlaAlaSpIle 285  
DB 183 AAAAGGAAAAAGAGAGGATTTAGAAAGGCGCAATCTGTCTTCAGCAGCAGCATTT 242

QY 286 AAlaAlaThrValGlnHIsThrMetAlaCySHIsleuValLyGArgThHIsArgAlaile 305  
DB 243 GGTGCACAGTACACACACATGCAATGCAATCTTGTGAAAAAGAACATCGGCTATT 302  
QY 306 LeuPheCyGlyGlnArgAspLeuPuroGlnaPnaPnaAlaValLeuValAlaSerGly 325  
DB 303 CGTITTTTGACAGAGACACTGTGTACTCTCAAAATTAAGCAAGTACTGTGATCTGGT 362  
QY 326 GlyValAlaSerAspPheThrIleArgArgAlaLeuGluIleleuThrAsnAlaThrGln 345  
DB 363 GGATGTGAGAGTAATCTTATATTCGAGAGCTCTGGAAATTTTAAACGACACAG 422  
QY 346 CysThrLeuLeuCySPoRProProArgLeuCySThrAspAsnGlylleMetIleAATP 365  
DB 423 TCCACTTTGTGTGTCCTCTCCACAGTATGCACTGATTAATGATTAATGATGATG 482  
QY 366 AAsnGlyTlleGluArgLeuArgAlaGlyLeuGlylleleuHIsAspIleGluGlylleArg 385  
DB 483 AATGTATATGAAGAAGTACGTCTGTGGCAATTTTACATGACATAGAAAGGCAATCGC 542  
QY 386 TyrGluProLyCySPoRLeuGlyValAspIleSerLySGlyValGlyAlaSerIle 405  
DB 543 TATGAACCAAAATGTCCTTGTGAGTACATATCAAAAGAACTTGAGAAAGCTTCATA 602  
QY 406 LysValProGlnLeuLySmetGluile 414  
DB 603 AAAGTACCAATTAATAATGAGATA 629

RESULT 13  
CN823245 730 bp mRNA linear EST 02-JUN-2004  
LOCUS  
DEFINITION  
Oa splbn\_04N08 M13reverse Sheep spleen\brain pSPORT1 library Ovis  
arles cDNA clone Oa\_splbn\_04N08 5', mRNA sequence.  
ACCESSION  
CN823245  
VERSION  
CN823245.1 GI:47951314  
KEYWORDS  
EST.  
SOURCE  
Ovis arles (sheep)  
ORGANISM  
Ovis arles  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Caprinae; Ovis.  
1 (bases 1 to 730)  
Gossner, A. and Hopkins, J.  
Ovine spleen\brain cDNA library  
Unpublished (2004)  
COMMENT  
Contact: J Hopkins  
Veterinary Biomedical Sciences  
University of Edinburgh  
Summerhall Square, Edinburgh, EH9 1QH.  
Email: j.hopkins@ed.ac.uk  
Plate: 04 Row: N Column: 08  
Seq primer: M13reverse  
High quality sequence start: 6  
High quality sequence stop: 550.  
Location/Qualifiers  
1. 730  
/organism="Ovis arles"  
/mol\_type="mRNA"  
/db\_xref="taxon:9940"  
/clone="Oa splbn\_04N08"  
/clone\_lib="Sheep spleen\brain pSPORT1 library"  
/note="Vector: pSPORT1"

ORIGIN  
Alignment Scores:  
Pred. No.: 1.08e-106 Length: 730  
Score: 1045.00 Matches: 204  
Percent Similarity: 93.51% Conservative: 12  
Best Local Similarity: 88.31% Mismatches: 15  
Query Match: 49.18% Indels: 0  
DB: 7 Gaps: 0

US-10-649-273-2 (1-414) x CN823245 (1-730)

```
Qy 96 GlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThr 115
Db 37 CAAGAAGCTCTCTCGCAGTAAGCTCTCCAAAGTAAGTCTCGACCAATTCGAACATACC 96
Qy 116 IleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuVal 135
Db 97 GTAAGCCAGGACTTGCTTAAAGCTTGGCGTAGAGTTATTCATTAGCTTACCAACGAGTA 156
Qy 136 GlyGlnLeuLysLysProPheIleProIleHisIsmetGluAlaHisAlaLeuThrIle 155
Db 157 GACCAAGTTAAAGCCCTTCATTCCTCCATTCATCATGAGGCTCATGCACTTACTATT 216
Qy 156 ArgLeuThrAsnLysValGlnPheProPheLeuValLeuLeuIleSerGlyGlyHisCys 175
Db 217 AGGTTAAACAATAAGTAGTAATTCATTTAGTTCCTTTGATTTCTGAGAGCTCATCT 276
Qy 176 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 195
Db 277 CTTTGGCATTAGCTAGAGAGATTTCAGATTTCTTCTTGGAAAGCTTTGACATA 336
Qy 196 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 215
Db 337 GCACCAAGCCGACATCTTGAAGAAGTAGCAAGACCTTGCTTAAATAAACAATCCAGAG 396
Qy 216 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 235
Db 397 TGCTCCACCATGAGTGGCGGAGGCTATAGAAACATTTGGCCAAACAGGAAATGATTG 456
Qy 236 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerThrGly 255
Db 457 CATTTTGATTTCCAACTCCCAACGACGCTTAAATAATTTGATTTTCTTTCTCGA 516
Qy 256 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGluLys 275
Db 517 CTTCAACATGTTATTTGATTAAGTAGTAAGTAAGCAAGAAAGAAAGAGTAATTCAGCAG 576
Qy 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 295
Db 577 GGGAGATCTCTGTCTTCAGCTGCACACATGCTGTCGACATCCAGACACCTTAGCTTC 636
Qy 296 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 315
Db 637 CACGTTGCAAAAGAACACATGCGCTATTTCTGCTGCAACAGAGAGCTTGTATCAT 696
Qy 316 GlnAsnAsnAlaValaLeuValAlaSerGlyGly 326
Db 697 CAAGTAACGCACTGCTGTATCTGAGGC 729

RESULT 14
CK833139 661 bp mRNA linear EST 04-MAR-2004
LOCUS DEFINITION 4057170 BARC 8BOV Bos taurus CDNA clone 8BOV_20020 5', mRNA
ACCESSION CK833139
VERSION CK833139
KEYWORDS EST. GI:45061881
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
1 (bases 1 to 661)
Baumann,R.G., Baldwin,R.L., Sonstegard,T.S., Van Tassel,C.P. and
Metukumalli,L.K.
Construction and Analysis of a cDNA Library Generated From
Intestinal Muscle and Epithelial Tissues of Holstein Cattle
Unpublished (2004)
Contact: Richard G. Baumann
Bovine Functional Genomics Lab
ARRI
BLDG 162: BARC-EAST, Beltsville, MD 20705, USA
Tel: 3015048604
```

```

Fax: 3015048744
Email: rbaumann@arri.barc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
0.000925 using options -trim alt ' ' -trim fasta. Vector identified
by cross match using options -mismatch 12 -minscore 18
Plate: 20 row: O column: 20
Seq primer: CTTATTTAGGTGACACTATAGAAC
High quality sequence stop: 661.
Location/Qualifiers
1. 661
/organism="Bos taurus"
/mol_type="mRNA"
/strain="Holstein"
/db_xref="taxon:9913"
/clone="8BOV 20020"
/sex="Female"
/tissue_type="Epithelial, Muscle"
/dev_stage="Lactating, Neonatal"
/lab_host="DH10B TONA"
/clone_1ib="BARC 8BOV"
/notes="Organ: Intestine; Vector: pCMVSPORT6.1; Site 1:
NOCI; Site 2: EcorI; Normalized cow CDNA intestinal
library in pCMVSPORT6.1, constructed from equimolar mRNA
pools derived from 5 sources, 4 lactating intestinal, 1
neonatal intestinal 4/5 Lactating, Proximal Duodenum,
Jejunum, Distal Ileum, Colon, 1/5 Neonatal, Proximal
Duodenum, Jejunum, Distal Ileum"

ORIGIN
Alignment Scores:
Pred. No.: 4,98e-106 Length: 661
Score: 1038.50 Matches: 203
Percent Similarity: 95.48% Conservative: 8
Beet Local Similarity: 91.86% Mismatches: 9
Query Match: 48.87% Indels: 1
DB: 7 Gaps: 1

US-10-649-273-2 (1-414) x CK833139 (1-661)
Qy 30 ProGlyThrLeuPheLeuHisLysIleValLeuGlyIleGluThrSerCysAspAspThr 49
Db 2 CCCGGG---ATATTCCTTCAATAACAGTATTTGGGAATTAACCACTTGATGATACAA 58
Qy 50 AlaAlaAlaValaLysProGluThrGlyAsnValLeuGlnGlyGlnAlaIleHisSerGlnThr 69
Db 59 GCACTGCTGTAGTGAAGTGAAGCTGGAATGTTTGGAGAGAGCAATACATTCACAACT 118
Qy 70 GluValHisLeuLysThrGlyGlyIleValProProAlaIleGlnGlnLeuHisArgGlu 89
Db 119 GAAGTTCATTTAAAGAACAGGTGGATTATTCCTCCAGTAGCTCAACGCTTCAATAGAA 178
Qy 90 AsnIleGlnArgIleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeu 109
Db 179 AATATTCACGCACTATACAGAAAGCTCTCTGCAAGTAGTAAGTCTCTCCAAAGTCAATC 238
Qy 110 SerAlaIleAlaThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSer 129
Db 239 TCAGCAGTTGCAACCAACCAATAAAGCCAGACTTGAAGCTTGGCGCTAGAGTTATGCG 298
Qy 130 PheSerLeuGlnLeuValGlyGlnLeuLysLysProPheIleProIleHisIsmetGlu 149
Db 299 TTTAGCTTCAACACTGTATAGCAACGATTTAAAGCCCTTCATTCCTCATTCACATGAG 358
Qy 150 AlaHisAlaLeuThrIleArgLeuThrAsnLysValGlnPheProPheLeuValLeuLeu 169
Db 359 GCTCAGCAGCTACATATAGTTAACAATAAGTGAATTCGTTTATGTTCTTTG 418
Qy 170 IleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeu 189
Db 419 ATTTGAGAGGTCATTTCTTTGGCATTAGTAGAGGAGTTCAGATTTTCTTCTTCTT 478
Qy 190 GlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSer 209
```

Db 479 GGGAAAGCTTTGACATGACACAGTGACATGCTTGAACAAGTAGCAAAAGACTTCC 538  
 Qy 210 LeuileuysHSPproGluCysSerThreSerGlyGlyValaIIeGluHisleuIa 229  
 Db 539 TTAATTAACATCCAGTGTCTCCACCATGATGCGGGAAGCTATAGAACATTTGGCC 598  
 Qy 230 LyeGInGlyAsnAcGPhelHisPheAspIleuysProProleuHisHISAlaIysAsnCys 249  
 Db 599 AAACAGGGAATAGATTGCAATTGATTGATTCACGCTCCATGCAACGTCCTAAATTTGT 658  
 Qy 250 Asp 250  
 Db 659 GAT 661  
 RESULT 15  
 BQ433135 879 bp mRNA linear EST 24-MAY-2002  
 LOCUS BQ433135  
 DEFINITION AGENCOURT 7760756 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:6016098  
 5', mRNA sequence.  
 ACCESSION BQ433135  
 VERSION BQ433135.1 GI:21172211  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 879)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LIML at:  
 http://image.llnl.gov  
 Plate: LIML3213 row: d column: 19  
 High quality sequence stop: 674.  
 FEATURES  
 source location/Qualifiers  
 1..879  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone\_image:6016098"  
 /issue\_type="embryonal carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_92"  
 /note="Organ: testis; Vector: PCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 2.5 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."  
 ORIGIN  
 Alignment Scores: 1.39e-103 length: 879  
 Pred. No.: 1018.50 Matches: 210  
 Score: 94.67% Conservative: 3  
 Percent Similarity: 93.33% Mismatches: 9  
 Best Local Similarity: 47.93% Indels: 3  
 Query Match: 5 Gaps: 1  
 DB: 5  
 US-10-649-273-2 (1-414) x BQ433135 (1-879)  
 Qy 1 MetleuileuThrystrThralaGlyValPhePheIysProSerIysArgIysValTyr 20  
 Db 201 ATGCTATATCTTGACTAAGACTGCAAGAGTTTCTTTTAAACATCAAAAAGAAAGTTTAT 260  
 Qy 21 GlupheIuysSerPheAsnPhelHisPheGlyThreIuPheIuHisIysIleValIeu 40

Db 261 GAATTTTAAGAAGTTTAAATTTTCATCTCGAACACTATTCTTCTATAAATAGTATG 320  
 Qy 41 GIYIleGluThSerCysAspAspThrAlaAlaValaIAspGluThrgIyAsnVal 60  
 Db 321 GGAATTTGAACATGATTGTGATATATACAGACGCTGCTGTGTGATGAATGAATGTC 380  
 Qy 61 LeuGlyGluAlaIleHisSerGlnThrgIyValaHisleuIysThrgIyGlyIleValPro 80  
 Db 381 TTGGGGAAGCAATACATTCCTCCAACTGAAGTTCAATTTAAAAACAGGTGGGATTTGCTT 440  
 Qy 81 ProAlaIleGInGluIleuHisArgGluAenIIeGlnArgIleValGInGluAlaSer 100  
 Db 441 CCAGCAGCTCAACAGCTTCACAGAGAAATATTTCACGATAGTACAGAGGCTCTTTCT 500  
 Qy 101 AlaserGlyValSerProSerAspIleuSerAlaIleAlaThrThrlleuysProGlyIleu 120  
 Db 501 GCCAGTGAAGTCTCTCCAGATGACCTCTCCAGCATTTGCAATGCAATTAACAGAGACTT 560  
 Qy 121 AlaIeSerIeuGlyValGlyIeuserPheSerIeuGlnIeuValGlyGlnIeuIysIys 140  
 Db 561 GCTTTAAGCTGGGAGCTGGGCTTATCATTTAGCTTACAGCTGGTAGGACAGTTAAAAAG 620  
 Qy 141 ProPheIleproIleHisIleGluAlaHisAlaIeuThrlleargIeuThraIys 160  
 Db 621 CCATTCATTCCTCATCATATGAGAGGCTCATGACCTTATAGTTGACCAATAA 680  
 Qy 161 ValGluPheProPheIeuValIeuIleuIleSerGlyGlyHisCysIeuIeuAlaIeuVal 180  
 Db 681 GTAGAAATTCCTTTTATGTTCTTTGATTTCTGAGGTCACGTCTGTGGCATTTGCT 740  
 Qy 181 GInGlyValSerAspPheIeuIeuIeuGlyIysSerIeuAspIleAlaProGlyAspMet 200  
 Db 741 CAAGAGCTTCACATTTTCCTGCTTCTTGAAAGTCTTTGACATAGCACAGGATGACATG 800  
 Qy 201 -IeuAspIysValAlaArgArgIeuSerIeuIle---LysHis-ProGluCysSerThrm 219  
 Db 801 GCTTGACAAAGTGGGCAAGAAAGACTTCTTTAATANAACATTCCAAGTGTCTCCCC 860  
 Qy 219 etSerGlyGly 222  
 Db 861 ATGGAANGGT 871

Search completed: November 11, 2005, 02:14:03  
 Job time : 5885 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 10, 2005, 16:54:39 ; Search time 293.85 Seconds  
(without alignments)  
2305.321 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125

Sequence: 1

BLASTSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n\_model -DEV=xlp  
-O=/cgn2\_1/USPRO\_spool\_p/US10649273/rnat\_02112005\_091339\_15581/app\_query.fasta\_1.1429  
-DB=Issued\_Patents\_NA -QWRT=fastcap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blowsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORF=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273 @CGN 1.1 177 @rnat\_02112005\_091339\_15581 -NCPU=6 -ICPU=3  
-NO MMAP -LARGESQURY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Issued Patents NA:\*  
2: /cgn2\_6/p2nata/1/ina/5A\_COMB.seq:\*  
3: /cgn2\_6/p2nata/1/ina/5B\_COMB.seq:\*  
4: /cgn2\_6/p2nata/1/ina/6A\_COMB.seq:\*  
5: /cgn2\_6/p2nata/1/ina/6B\_COMB.seq:\*  
6: /cgn2\_6/p2nata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	2125	100.0	2197	4	US-10-067-443-1
2	2090.5	98.4	1387	4	US-10-067-443-21
3	1747	82.2	1416	4	US-09-774-528-177
4	1385	65.2	14364	4	US-10-067-443-23
5	1186.5	55.8	14364	4	US-10-067-443-20
6	502	23.6	94750	4	US-09-596-002-38
7	501	23.6	1053	4	US-09-540-236-806
8	499.5	23.5	1206	4	US-09-252-991A-801
9	432.5	23.2	1059	4	US-09-252-991A-884
10	482.5	22.7	1074	4	US-09-543-681A-2341
11	482	22.7	996	4	US-09-902-540-6612
12	482	22.7	2582	4	US-09-902-540-503

13	479	22.5	1315	1	US-08-087-797-1	Sequence 1, Appl1
14	475.5	22.4	1830121	4	US-09-557-884-1	Sequence 1, Appl1
15	475.5	22.4	1830121	4	US-09-643-990A-1	Sequence 1, Appl1
16	474.5	22.3	1032	4	US-09-489-039A-2050	Sequence 2050, Ap
17	472	22.2	1092	4	US-09-107-532A-2955	Sequence 2955, Ap
18	461	21.7	1008	3	US-08-987-121A-5	Sequence 5, Appl1
19	461	21.7	1011	3	US-08-987-121A-3	Sequence 3, Appl1
20	457	21.5	1006	3	US-08-961-083-51	Sequence 51, Appl1
21	457	21.5	1006	3	US-09-536-784-51	Sequence 51, Appl1
22	457	21.5	10974	3	US-08-961-527-214	Sequence 214, App
23	455	21.4	1011	3	US-09-066-512-1	Sequence 1, Appl1
24	454	21.4	1101	4	US-09-134-000C-1551	Sequence 1551, Ap
25	453	21.3	1011	4	US-09-583-110-2196	Sequence 2196, Ap
26	448	21.1	1011	4	US-09-107-433-1618	Sequence 1618, Ap
27	443	20.8	1107	3	US-09-134-001C-1072	Sequence 1072, Ap
28	433	20.4	1026	3	US-09-149-624-1	Sequence 1, Appl1
29	424	20.0	15249	4	US-08-966-171E-102	Sequence 102, App
30	424	20.0	15249	4	US-08-781-868A-102	Sequence 102, App
31	406.5	19.1	640681	4	US-09-790-988-1	Sequence 1, Appl1
32	406.5	19.1	1230025	4	US-09-198-452A-1	Sequence 1, Appl1
33	406.5	19.1	1230025	4	US-09-438-185A-1	Sequence 1, Appl1
34	404.5	19.0	1155	4	US-09-602-777A-147	Sequence 147, App
35	404	19.0	3064	3	US-09-221-017B-794	Sequence 794, App
36	379.5	17.9	4403765	3	US-09-103-840A-2	Sequence 2, Appl1
37	379.5	17.9	4411529	3	US-09-103-840A-1	Sequence 1, Appl1
38	369.5	17.4	645	4	US-09-328-352-483	Sequence 483, App
39	362	17.0	876	4	US-09-724-623-19	Sequence 19, Appl1
40	357	16.8	580073	4	US-08-545-528D-1	Sequence 1, Appl1
41	333.5	15.7	42325	4	US-08-311-731A-131	Sequence 131, App
42	327.5	15.4	3993	4	US-09-710-279-3985	Sequence 3985, Ap
43	319	15.0	822	4	US-09-710-279-727	Sequence 727, App
44	317	14.9	36941	4	US-08-311-731A-130	Sequence 130, App
45	313	14.7	1664976	4	US-08-916-421B-1	Sequence 1, Appl1

#### ALIGNMENTS

RESULT 1  
US-10-067-443-1  
Sequence 1, Application US/10067443  
Patent No. 6642041  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
PRIOR FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 2197  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (231) ... (1472)  
US-10-067-443-1  
Alignment Scores:  
Pred. No.: 5,72e-261  
Score: 2125.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
DB: 4  
Gaps: 0  
US-10-649-273-2 (1-414) x US-10-067-443-1 (1-2197)  
Length: 2197  
Matches: 414  
Conservative: 0  
Mismatches: 0  
Indels: 0  
Gaps: 0



QY 1 MetLeu1leuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
Db 231 ATGCTAACTTACATAGAGCTCAGAGAGTTTAAAAACCATCAAAAAGAAAGTTAT 290  
QY 21 GluPheLysArgSerPheAsnPheHisProGlyThrLeuPheLysIleValLeu 40  
Db 291 GAATTTTAAAGAAAGTTTAAATTTTCACTCGAACAACATAATTTCTCATAAATAGATATG 350  
QY 41 GlyIleGlyThrSerCysAspAspThrAlaAlaValValAspGlyThrGlyAsnVal 60  
Db 351 GGAATTTGAAACATGATGATGATATACAGACGCTGCTGATGATTAACCTGAAAATGTG 410  
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80  
Db 411 TTGGAGAAAGCAATACATTCCTCCAACTGAAGTTCACTTAAAAACAGTGGAGATTTCTCT 470  
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
Db 471 CCAGACGCTCAACAGCTTCACAGAGAAATATTCAACGAATAGTACAGAGAGCTTTCTCT 530  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
Db 531 GCCAGTGAAGTCTCTCCAGAGTACCTCTCAGCAATTCGCAACTAACCAAGAGACTT 590  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140  
Db 591 GCTTTAAGCTCGGAGTGGCTTTATCATTTAGCTTACAGCTGGTAGACAGTTTAAAAAG 650  
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
Db 651 CCATTCAATCCCATTCATATAGAGGCTCATAGCACTTACATTAAGTTGATTGACCAATAAA 710  
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
Db 711 GTAGAAATTTCTTTTATTTTACTTTCTTTTGTGAGGTCACATGCTGTGTGGCATTAAGT 770  
QY 181 GlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
Db 771 CAAAGAGTTTCAGATTTTCTGCTTCTTGAGAACTTTTGACATAGCACAGGTTGACATG 830  
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyLysSerThrMetSer 220  
Db 831 CTTCAGCAAGGTGGCAAGAGACCTTTCTTTATTAACATCCAGAGGCTTCACACAGAGT 890  
QY 221 GlyIleLysValAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
Db 891 GGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATTTCAATTTTGACATCAAA 950  
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
Db 951 CTTCCCTTCATCAGTACGTAATAAATGTGATTTTCTTTTACTGAACTTCAACACGTTACT 1010  
QY 261 AspLysIleIleMetCysLysGlyLysGlnGluGlyIleGlnLysGlyGlnIleLeuSer 280  
Db 1011 GATTAATAATATATGAAAAAGAAAAAGAAAGATTTGAGAAAGGCAAAATCTCTCT 1070  
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisSleuValLysArg 300  
Db 1071 TCAGCAGCAGACATTCCTCCACAGTACAGCACACATGCGCATCTTGTGAAAAAGA 1130  
QY 301 ThrHisArgValAlaLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320  
Db 1131 AACACATCGGGCTAATTTGTTTGTAAAGCAGAGAGACTTTTACTCAAAATTAATGCGTA 1190  
QY 321 LeuValAlaSerGlyValValAlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeu 340  
Db 1191 CTGGTTCATCTGAGTGTGCTCGCAAGTAATCTTAATCCGACAGCTCTGGAATTTTA 1250  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360  
Db 1251 ACAAAGCAACACAGTGCATTTGTTGTGTCTCTCCACAGTATGCACTGATTAATAGGC 1310  
QY 361 IleMetIleAlaThrAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380

Db 1311 ATTATGATTCAGAGATGATATGAAAGCTACGTCTGCTGGCATTTTACATGAC 1370  
QY 381 IleGluGlyIleArgIleGlyProLysCysProLeuGlyValAspIleSerLysGluVal 400  
Db 1371 ATTAGAAGGCAATCGCTATGAAACAAATATGCTCTTGAGTAAACATATCAAAAGAGTT 1430  
QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
Db 1431 GGAAGAGCTTCATTAAGATACCAATTAATAATGAGATA 1472  
RESULT 2  
US-10-067-443-21  
; Sequence 21, Application US/10067443  
; Patent No. 6642041  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED  
; TITLE OF INVENTION: SPINAL CORD, MP-1  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 21  
; LENGTH: 1387  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-067-443-21  
Alignment Scores:  
Pred. No.: 6,62e-257 Length: 1387  
Score: 2090.50 Matches: 412  
Percent Similarity: 93.85% Conservative: 0  
Best Local Similarity: 93.85% Mismatches: 2  
Query Match: 98.38% Indels: 25  
DB: 4 Gaps: 1  
US-10-649-273-2 (1-414) x US-10-067-443-21 (1-1387)  
QY 1 MetLeu1leuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
Db 24 ATGCTAACTTACATAGAGCTCAGAGAGTTTAAAAACCATCAAAAAGAAAGTTAT 83  
QY 21 GluPheLysArgSerPheAsnPheHisProGlyThrLeuPheLysIleValLeu 40  
Db 84 GAATTTTAAAGAAAGTTTAAATTTTCACTCGAACAACATAATTTCTCATAAATAGATATG 143  
QY 41 GlyIleGlyThrSerCysAspAspThrAlaAlaValValAspGlyThrGlyAsnVal 60  
Db 144 GGAATTTGAAACATGATGATGATATACAGACGCTGCTGATGATTAACCTGAAAATGTG 203  
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80  
Db 204 TTGGAGAAAGCAATACATTCCTCCAACTGAAGTTCAATTAATAAGGATGGAATGTCTCT 263  
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
Db 264 CCAGACGCTCAACAGCTTCACAGAGAAATATTCAACGAATAGTACAGAGAGCTTTCTCT 323  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
Db 324 GCCAGTGAAGTCTCTCCAGAGTACCTCCAGCAATTCGCACTTACATTAACCAAGAGACTT 383  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140  
Db 384 GCTTTAAGCTCGGAGTGGCTTTATCATTTAGCTTACAGCTGGTAGACAGTTTAAAAAG 443  
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160



Db 880 GAGAGGGGCAAAATCTGCTTTCAGACAGACATTCGTCACAGTACAGACACAAATG 939  
Qy 294 A|A|C|Y|S|I|S|L|E|U|V|L|V|A|S|I|G|T|H|S|A|I|L|E|U|P|H|E|C|Y|A|L|E|S|I|N|A|S|E|P|H|E|U 313  
Db 940 GCAATGCTCTTGTGTAAGAAACACATGCGGCTATTCGTGTTGTGTAAGAGAGCTTG 999  
Qy 314 LeuProGlnAAsnAlaValLeuValAlaSerGlyValAlaSerAsnPhetYrile 333  
Db 1000 TTAACTCAAAATATAGCACTACTGCTTCATCTGCTGGTGGCAAGTAACTTATATAC 1059  
Qy 334 ArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProPro 353  
Db 1060 CGCAGAGCTCTGGAATTTTAAACCAACCAACAGTGCATTTGTTGCTCTCTCC 1119  
Qy 354 ArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGlnArgLeuAla 373  
Db 1120 AGACTATCCTGGAATGAGCTTATGATGATGATGATGATGATGATGATGATGATGAT 1179  
Qy 374 G|L|E|U|G|I|L|E|U|H|I|S|A|P|I|L|E|G|I|L|E|A|T|G|Y|G|I|U|P|R|O|L|Y|S|C|Y|S|P|R|O|L|E|U|G|I| 393  
Db 1180 GCGTTGGCATTTCATGACATAGAGGATCCGCTATGAAACCAAAATGCTCTTGA 1239  
Qy 394 ValAspIleSerLysGluValGlyValaSerIleLysValProGlnLeuLysMetGlu 413  
Db 1240 GTAGCATATCAAAAGAGTGGAGAGCTTCATTAAGTACCAATTAATAATGAG 1299  
Qy 414 I|L|E|A|I| 414  
Db 1300 ATA 1302

RESULT 4  
US-10-667-443-23  
; Sequence 23, Application US/10067443  
; Patent No. 6642041  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN  
; TITLE OF INVENTION: SPINAL CORD, MP-1  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 23  
; LENGTH: 1526  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-667-443-23

Alignment Scores:  
Pred. No.: 1 62e-166 Length: 1526  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 65.18% Indels: 0  
DB: 4 Gaps: 0

US-10-649-273-2 (1-414) x US-10-667-443-23 (1-1526)

Qy 148 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 167  
Db 1 ATGGAGGCTCATGCACTTACTATAGTGTACCAATTAAGTAAATTTCTTTTATAGT 60  
Qy 168 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187  
Db 61 CTTTGATTTTCGAGAGTCACTGCTGTGTCATTTGATTCAGAGAGTTTTCAGATTTTCTG 120  
Qy 188 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 207

Db 121 CTTCTGGAAGTCTTTGACATAGACCAAGTGAATGCTTTACCAAGGTGCAAGAGA 180  
Qy 208 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysAlaIleGluHis 227  
Db 181 CTTTCTTTAATTAATCAATCAGAGTCTCCACCATAGTGTGGAAAGCATATAGACAT 240  
Qy 228 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 247  
Db 241 TTGGCCAAACCAAGAAATAGATTTCATTTGATCATCAAACTCCCTTGCACTCATGCTAAA 300  
Qy 248 AsnCysAspPheSerThrThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 267  
Db 301 AATTTGATTTTCTTTTACTGACCTTCAACAGTTCATATATAATTAATTAAGAAAAAG 360  
Qy 268 G|L|U|Y|G|I|U|G|I|L|E|U|L|Y|G|I|N|I|L|E|U|S|E|S|E|A|A|A|S|E|P|H|E|A|A| 287  
Db 361 GAAAAAGAGAGATTTTGAAGAGGGCAAAATCTGCTTCAGACAGACATTTGCTGCC 420  
Qy 288 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 307  
Db 421 ACAGTACAGACACAATGAGATGTCATCTTGAAAAAGAACACATCGGCTATTCTGTTT 480  
Qy 308 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 327  
Db 481 TGTAAACAGAGACATCTGTAACTCAAAATTAATGCAAGTACTGTTGATGATGATGATGAT 540  
Qy 328 AlaSerAsnPhetYrileArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 347  
Db 541 GCAAGTACTTCTATATCCCAAGAGCTTGGAATTTTAACCAAGCAACAGTGCAT 600  
Qy 348 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 367  
Db 601 TTGTTGTGCTCTCTCCCAACATATGACATGATATGATGATGATGATGATGATGATGAT 660  
Qy 368 I|L|G|I|U|A|G|L|U|A|G|I|L|E|U|H|I|S|A|P|I|L|E|G|I|L|E|A|T|G|Y|G|I|U| 387  
Db 661 ATTTGAAGACTTACGCTGCTGCGCATTTTACATACATAGAGCATCCGCTATGAA 720  
Qy 388 ProLysCysProLeuGlyValAspIleSerLysGluValGlyValaSerIleLysVal 407  
Db 721 CCAAAATGCTCTTGGAGTACATATCAAAAGAGTGGAGAGCTTCATTAAGTA 780  
Qy 408 ProGlnLeuLysMetGluIle 414  
Db 781 CCACATTAATAATGAGATA 801

RESULT 5  
US-10-667-443-20  
; Sequence 20, Application US/10067443  
; Patent No. 6642041  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN  
; TITLE OF INVENTION: SPINAL CORD, MP-1  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 20  
; LENGTH: 14364  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-667-443-20

Alignment Scores:  
Pred. No.: 2 19e-139 Length: 14364  
Score: 1186.50 Matches: 313  
Percent Similarity: 32.30% Conservative: 1

Best Local Similarity: 32.20% Mismatches: 2  
Query Match: 55.84% Indels: 657  
DB: 4 Gaps: 4  
US-10-649-273-2 (1-414) x US-10-067-443-20 (1-14364)

QY 74 LysThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArg 93  
DB 10623 AGAACAGGTGGATGTTGCTTCACAGCTCAACAGCTTCACAGAAAAATATTCACAG 10682

QY 94 ILeValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113  
DB 10683 ATAGACAAAGAGCTCTTCTCCAGTGAGTCTCTCCAGTGACCTCTCAGCAATTCGA 10742

QY 114 ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133  
DB 10743 ACTACCATRAAACCGAGCTTCTTAACTGGGAGTGGGCTTATCATTTAGCTTACAG 10802

QY 134 LeuValGlyGlnLeuLysLysProPheIleProIleHisHisMetGluAlaHisAlaLeu 153  
DB 10803 CTGGTAGACAGTTAAAAAGCATTCATCCCATTCATCATATGAGGCTCATGCACTT 10862

QY 154 ThrIleArgLeuThrAspLysValGluPheProPheLeuValLeuLeuIleSerGlyGly 173  
DB 10863 ACTATAGGTGACCAATRAAGATAGCAATTCCTTTTATGTTCTTTGAATTCGAGAGT 10922

QY 174 HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeu 193  
DB 10923 CACTGCTGTTGGCATTTAGTTCAAGAGTTTCAGATTTTCTGCTTCATGGAAGTCTTGG 10982

QY 194 AspIleAlaProGlyAspMetLeuAspLysVal----- 204  
DB 10983 GACATAGACACGAGTGACATGCTTGCAAGGT-AAATTAAGAAATTAATTCATTCCTT 11041

QY 204 ----- 204  
DB 11042 TTGTGTATGTTGTCATTCACTAAGACATAGATGTGTACCACTTCACTTAA 11101

QY 204 ----- 204  
DB 11102 TATTTTCGAATTTTATCTTAGTAACTGAAAAAATTCACATATGTGTAGAAAAATAGA 11161

QY 204 ----- 204  
DB 11162 AAGAGTAGTACACATTTTATTAATCTTAGCCTTCTTAATAAATGTGAAGTTCAT 11221

QY 204 ----- 204  
DB 11222 ATCTGTACATAAGGCTGAAAAATAGTTTGACATACAGTTATGTATTTGCCAATATATGT 11281

QY 204 ----- 204  
DB 11282 ATGTGAAGAACTGCTTGTAACTAACTACATCTGCAAAAAAGTAAATAAGAGATAT 11341

QY 204 ----- 204  
DB 11342 ATATAGATTAACATAGGACATTAAGATGCAATGACAGATTAATTCACACATTTACT 11401

QY 204 ----- 204  
DB 11402 TACACACAGACAGGCTCCCCCACTTGTGTTTAAATATAGTACAGAGGCTACT 11461

QY 204 ----- 204  
DB 11462 GCCATATATAGAAAAACAACAACAACAACAACAAGCTTCCACAGTGAATTA 11521

QY 204 ----- 204  
DB 11522 TAGAGATATAGACAGATGTTCTTATTTATGAGTTCAATCATTAAGCAGTTATTTGCACT 11581

QY 204 ----- 204  
DB 11582 TCAGCCCATTTTCCAACTAATAGAGAAAGCAATAGACAGGGGCGAGTGTGCTCT 11641

QY 204 ----- 204  
DB 11642 TATGTTGGGTATCATTAAGAACAGAGTTGTCTGTACTGATTAATATGACTATAGTGC 11701

QY 204 ----- 204  
DB 11702 TATATTTGCCAAGTATAGCATGTTTATTCATTCAGGGGTTTTTGTGTTATGTAAT 11761

QY 204 ----- 204  
DB 11762 TTTCAATTTATTTCTTGGCATCTTTTGTTCACAGATTTTAAATTTATGACTTAA 11821

QY 205 -----AlaArgArgLeuSerLeuIleLysHisProGluCysSer 217  
DB 11822 AATATGTTCTTGTATAGTGGGAAGAGCTTTCTTAATTAACAATCCAGAGTGTCTCC 11881

QY 218 ThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPhe 237  
DB 11882 ACCATGAGTGTGGGAAGCATAGAACATTTGGCCAAACAGGAATAGATTTCAATTT 11941

QY 238 AspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGln 257  
DB 11942 GACATCAAACTCTCCCTTGACATGCTTAAATAATGTGATTTTCTTTTACTGACCTCAA 12001

QY 258 HisValThrAspLysIleIleMetLysLysGlyLysGlu----- 271  
DB 12002 CACTTACTGATTAATAATTAATGAAGAAAAAGAAAGAGATATTTCTTAATTAATGT 12061

QY 272 -----G1 272  
DB 12062 AAGTTGAACAGATTAATATTTCTCGATGTGCTTAAATAAGCTGCTCATTTCTGCAGG 12121

QY 272 YIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHis 292  
DB 12122 TATGGAAGGGGCAATCTCTGCTTCAGACAGACATTTGCTCCACAGTACAGCAC 12181

QY 292 rMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAs 312  
DB 12182 AATGCAATGTCATCTGTGAAAGAACACATCGGCTATTTCTGTTGTAAGCAGAGGA 12241

QY 312 PLeuLeuProGlnAsnAsnAlaValLeu----- 321  
DB 12242 CTTGTTACCTCAAAATATAGCATTACTGTGAAGTTTATGCTCATTTATAGTAATAGT 12301

QY 321 ----- 321  
DB 12302 CACTTGCATATATGTTACTTTTTCACAGACCTTGACTGTGTAGATGAACAGAT 12361

QY 321 ----- 321  
DB 12362 CTTTATGCTTATGTAAGCTGACAGTATGAAATTAAGCAGATAGAAAACTAACAG 12421

QY 321 ----- 321  
DB 12422 CCAATTTCTGTACTAGTTTGTAGCTTTAATGGAACAGCTGTATAGCTTCTTAAGCATA 12481

QY 321 ----- 321  
DB 12482 AGCTAATTTTGCATCTCTGTGTGATTTTAAAGAGGCTTACAAATTAAGAAATTAAT 12541

QY 321 ----- 321  
DB 12542 GCAGTAACGTATACATAATTTTAAAGAAATAGTGAGATTTCTTCATCTTTGATGA 12601

QY 321 ----- 321  
DB 12602 ATCCCTTGTGTTGTTGTTTATTAATTAAGCCAGTCAAAATTTAGCAGTGGAGGTGAT 12661

QY 321 ----- 321  
DB 12662 TCCAACTTTGTGACATAATATGTGATTAAGTTCTGATATTCAGCTATATTTGACAGCC 12721

```
Qy 321 ----- 321
Db 12722 AAAATCCCTTAATTGCTTAAAAAGCCTTGACAAACATCCGTTTAACGTAACTTAA 12781
Qy 321 ----- 321
Db 12782 CTTTATTCATTTAAAAATTATTAACATAAGTGGAAAAATGTTTAAATGTAATTCAT 12841
Qy 321 ----- 321
Db 12842 AGATGAATTTTACATGATATCAAGAATAATTTTTCAGAGTATAGTAAATGCA 12901
Qy 321 ----- 321
Db 12902 CAAATATATAAATTTTACGGCTTAAATAGTACTATGATGAAATTAATTAATA 12961
Qy 321 ----- 321
Db 12962 AATATTTAGATGAAGTGGAGAAATAATACAAAATGCTAGTAATGTTGATGCTA 13021
Qy 321 ----- 321
Db 13022 TTGAATATTAATTAATTTTCTTCCAAATTTTATACATAGATATGTCATGCTCC 13081
Qy 321 ----- 321
Db 13082 CATTACCATCTCAAAATGGAGTATTAATTAATGCTGATATTTTCTCCAGG 13141
Qy 321 ----- 321
Db 13142 TTTAATTAGCAGCTTGTCATATCCATATATAGTATTTGGTTTCTCAATTCCT 13201
Qy 322 ---ValAASerGlyIyValAASerAspNheTyrIleArgAlaLeuGluIleLe 340
Db 13202 TCAGGTCATCTGGTGGTGTGCAAGTAACTTATATATCCGACAGGCTCTGGAATTTT 13261
Qy 340 wThraAsnAlaThrGlnCysThrLeuLeuCysProPheArgLeuGluCysThrAspAsnG 360
Db 13262 AACAAACCCACACAGTGCATTTGTGTCTCTCCACAGATATGACATGATATATG 13321
Qy 360 YIleMetIleAlaTyr----- 365
Db 13322 CATTTATGATGATGATGTAAGCACAGATATACGTGCTTCACATCAATCATATTAAT 13381
Qy 365 ----- 365
Db 13382 TAATTGCATTTTATCATACTAAGCCTTCTCTTCAGATCTTGAGCTATTGATTTTAT 13441
Qy 366 -----AenGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuH 379
Db 13442 TTTAATGCTTCTTAATTTAGAAATGTATTTGAAAGACTACGCTGCTGGCATTTTTAC 13501
Qy 379 IAspIleGluGlyIleArgTyrGluProLys 389
Db 13502 ATGACATAGAAAGCATCCGCTATGAACCAAG 13533

RESULT 6
US-09-596-002-38
; Sequence 38, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Paterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US/09/596, 002
; CURRENT APPLICATION NUMBER: US/09/596, 002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 38
```

```
; LENGTH: 94750
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte template ID No. 6632636 38
; PUBLICATION INFORMATION:
US-09-596-002-38

Alignment Scores:
Pred. No.: 2,66-50 Length: 94750
Score: 502.00 Matches: 129
Percent Similarity: 50.81% Conservative: 59
Best Local Similarity: 34.86% Mismatches: 149
Query Match: 23,62% Indels: 34
DB: 4 Gaps: 9

US-10-649-273-2 (1-414) x US-09-596-002-38 (1-94750)
Qy 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThr--- 57
Db 23873 GATTTGGATTTGGAGACATCTTGTGATGAACAGGGCTTCCATCATGATAGCATG 23932
Qy 58 -----GlyAsnValLeuGlyGluAlaIleHisSerGlnThrGluValHisLeu 73
Db 23933 AATGGCGTGTGGCGGCGGCTGTCTCAGGTTTGTATAGCCAAATTAATTAATCATG 23992
Qy 74 LysThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArg 93
Db 23993 ACCATATGGGTGTCTGTGCTGAGCTTGCAGTGCAGTGCAGCATTCGTAAGCTTGTC 24052
Qy 94 IleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113
Db 24053 TATTTAATGAGCTGTGTGATCAGGCAATATACCAATCCGATTTGATGCGGTGCG 24112
Qy 114 ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133
Db 24113 TATACCAAGGGGCGCAGGGCTGATTTGGGGCGTTGATGACAGGGGATTAATTTGG 24172
Qy 134 LeuValGlyGlnLeuLysLysProPheIleProIleHisIleMetGluAlaHisAlaLeu 153
Db 24173 CTGGCGTATGGCGTGGCGGCTGCAGCGGTGTGGGTGATCATATGAGGAGCATCTGTT 24232
Qy 154 ThrIleArgLeuThr-----AsnLysValGluPheProPheLeuValLeuLeuIleSer 171
Db 24233 GCAACCGTATTTGGCCAGTATGACCCCAAGCTTTCGTTTGTGTGCTCTGCTGTG 24292
Qy 172 GlyIyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLys 191
Db 24293 GGGCGTATACCATGCTGTGCGCTGCCGAGTGTGTGGCGGTGATCATGATTTGGGCGAG 24352
Qy 192 SerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIle 211
Db 24353 TCTATCATGATGCGGTGGGTGATGCTTGTATTAACGCAAAAATGCTCAAACTG--- 24409
Qy 212 LysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGlyHisLeuAlaGln 231
Db 24410 CCTATCT-----GTTGGCCCAATATATGCAAAAATTTGCAAAAAC 24451
Qy 232 GlyAsnArgPheHisPheAspIleLysProProLeuHisIleAlaLysAsnCysAspPhe 251
Db 24452 GGCACCCACACCGCTTGTAGCTGCCCAAGACCCATGACAT---AAAGGCTGATTTT 24508
Qy 252 SerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysGlyLeuGluGlu 271
Db 24509 TCGTTGATGTCATGAAGAACCGCATTCATATATCTCAAAAGACACCAAGCCCA 24568
Qy 272 GlyIleGluLysGlyGlnIleLeuSerSerAlaAlaSerIleAlaIleThrValGlnHis 291
Db 24569 AGCAGACCC-----GCCACACGACGACGACATCGCGCAAGCTTTGAGATAT 24613
Qy 292 ThrMetAlaCysHisIleValLysArgThrHisArgAlaIleLeuPheCysLysGluArg 311
DB: 4 Indels: 34 Gaps: 9
```

Db 24614 GCGGTGTGGATCTTGTGCTCAAAAATGACCAAGCACTACAGATGACAGGCACTTCC 24673  
Qy 312 AspleuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsn 331  
Db 24674 CAG-----CTGTGTGTGCGAGGGGGCGCTCTGCGCAATCAG 24709  
Qy 332 TyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysPro 351  
Db 24710 ATGCTACGCGCGACCTGTACCGAGACGCTCCGCCAATGCATGCGGTACTATGCC 24769  
Qy 352 ProProArgLeuGlySerThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeu 371  
Db 24770 CCGACCGACCTATGACCGGATATGCTGCGATGATGCTGCTGCTGCTTGTCCGCTC 24829  
Qy 372 ArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysPro 391  
Db 24830 AGCTGTGGACAGTGC-----GATGACTTGGC-GGTTCCGTATTC-----CCG 24873  
Qy 392 LeuGlyValAspIleSerTyrGluValGly 401  
Db 24874 ATGGATATGACGACGCTTGGCGTATCGGC 24903  
RESULT 7  
US-09-540-236-806  
Sequence 806, Application US/09540236  
Patent No. 6673910  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR  
FILE REFERENCE: 2709.2005-001  
CURRENT APPLICATION NUMBER: US/09/540,236  
CURRENT FILING DATE: 2000-04-04  
NUMBER OF SEQ ID NOS: 3840  
SEQ ID NO 806  
LENGTH: 1053  
TYPE: DNA  
ORGANISM: M.catarrhalis  
US-09-540-236-806  
Alignment Scores:  
Pred. No.: 1.23e-53 Length: 1053  
Score: 501.00 Matches: 126  
Percent Similarity: 49.86% Conservative: 56  
Best Local Similarity: 34.52% Mismatches: 139  
Query Match: 23.58% Indels: 44  
Gaps: 8  
US-10-649-273-2 (1-414) x US-09-540-236-806 (1-1053)  
Qy 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaSerGluThr--- 57  
Db 16 GTATTGGATTGGACACTTGTGATGAACAAGGGCTTGCATCATATAGTACGATG 75  
Qy 58 -----GlyAsnValLeuGluIleValAlaIleHisSerGlnThrGluValHisLeuLys 74  
Db 76 AATGCGAGGGGGGGGCTGTCTGTCTGACGTTTATATAGCCAATATATATACATGCGACC 135  
Qy 75 ThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIle 94  
Db 136 TATGCGGCTGTGCTGCTGACCTGCGACGACCAATTCGTAAAGCTTGGCCGTTA 195  
Qy 95 ValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThr 114  
Db 196 TTTAATGAGATCTTGATCAAGCAATATACCAATTCGACATTCATGTCGGGTAT 255  
Qy 115 ThrIleLysProGlyIleValAlaLeuSerLeuGlyValGlyLeuSerSerLeuGlnLeu 134  
Db 256 ACCAAAGGCCCGGGCTATTTGGGCAATTGATGACAGGGGCAATTATTTGGCGGACCGTG 315  
Qy 135 ValGlyGlnLeuLysProPheIleProIleHisIleMetGluAlaHisAlaLeuThr 154  
Db 316 GCGTATGGGCTTGGCGGTGCGACGGTGGCGGTGATCATATGAGGAGACATCTTTAGCA 375

Qy 155 IleArgLeuThr-----AsnLysValGluPheProPheLeuValLeuLeuIleSerGly 172  
Db 376 CCGCTATTGGCCAGTGAATGACCCCCCAAGCTTTCGTTTGTGCTGCTGTGTGCGGCG 435  
Qy 173 GlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSer 192  
Db 436 GGTCTATCATGCTGTGGCGGTGCCGATGTGGCGGTGATTCAGATATTCGAGGAGGCTC 495  
Qy 193 LeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgGlyLeuSerLeuLys 212  
Db 496 ATCATATATGCGGGGCGGATGCTTTGATTAACGGCAAAATGCTCAACTG---CCC 552  
Qy 213 HisProGluCysSerThrMetSerGlyIleLysAlaIleGlnHisLeuAlaLysGlnGly 232  
Db 553 TATCCT-----GGTGTCCAAATATCGAAAATTAAGCCAAAGAGCG 594  
Qy 233 AsnArgPheHisPheAspIleLysProProLeuHisIleValLysAsnCysAspPheSer 252  
Db 595 AACCCACACCGCTATGAGCTGCCAAGACCGATGACGAT--AAAGGCTGGATTTTTCG 651  
Qy 253 PheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGluGluGly 272  
Db 652 TTCAATGGCATGAAAACCGCATTCATATCTCAAGACACGCCAAACGCCAAGC 711  
Qy 273 IleGluLysGlyIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnIleThr 292  
Db 712 GACCT-----GCCACACAGACACACTCCCGCACTTTGAGTATGCG 756  
Qy 293 MetAlaCysHisLeuValLysArgThrHisIleArgAlaIleLeuPheCysLysGlnArgAsp 312  
Db 757 GTGTGGATGATCTTTGTGCAAAAATGACCAAGGACCTACAGATGACAGGATTCGCCAG 816  
Qy 313 LeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAspPheTyr 332  
Db 817 -----CTGTGTGTGCGACGGGGCGCTCTGCGCAATCAGACG 852  
Qy 333 IleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProPro 352  
Db 853 CTACGCGGACCTGTACCGGACGCGCTGCCCAATGATGCGGTGCTACGCGCCCG 912  
Qy 353 ProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArg 372  
Db 913 ACCGATATGACCGGATATGATGATGATGATGCTTACCGCTTTGTCCGCTAACG 972  
Qy 373 AlaGly-----Leu 375  
Db 973 CGTGGCAGTCGATGACTTGGCGGTGCTGTATTCGCCAGATGATATGACGATGCTT 1032  
Qy 376 GlyIleLeuHisAsp 380  
Db 1033 GGTATCGAATATGAT 1047  
RESULT 8  
US-09-252-991A-801  
Sequence 801, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 801  
LENGTH: 1206  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa



US-09-252-991A-801

## Alignment Scores:

Pred. No.:	2,436-53	Length:	1206
Score:	499.50	Matches:	133
Percent Similarity:	51.68%	Conservative:	67
Best Local Similarity:	34.37%	Mismatches:	158
Query Match:	23.51%	Indels:	29
		Gaps:	10

US-10-649-273-2 (1-414) x US-09-252-991A-801 (1-1206)

```
Qy 13 LysProSerLysArgLysValTyrGluPheLeuArgSerPheAsnProGlyThr 32
Db 91 AAAGCTCCGATCGAATAAACAACCTGCTGCGCAAGCCCGCATTTATCATGCGGCTT 150
Qy 33 LeuPheLeuHisLysIle-----ValLeuGlyIleGluThrSerCys 46
Db 151 CGTATTGCGCAAGCTCTAAGGCTCAAGCCCATGCGCTGCGGACTGGAATAAGCTCTGC 210
Qy 47 AspAspThrAlaAlaAlaValAlaArgLysValAsnValLeuGlyValAlaIleHis 66
Db 211 GACGGAACCGCGCTGCGCTTACGACGACGAAACCGGCGCTGCGGACGCGCTGTC 270
Qy 67 SerGluThrGluValHisLeuLysThrGlyIleValProProAlaIleGluThr 86
Db 271 AGTCAGATGACCTTCATCGCGCTTACGCGGCGCTGCTGCGGCTGCGGCTGCGGCTG 330
Qy 87 HisArgGluAsnIleGluArgIleValGluGluAlaLeuSerLysArgIleValSerPro 106
Db 331 CACGTCAGAGCGCATGCTGCGCGGATCGCGGATCGCGGATCGCGGATCGCGGATCG 390
Qy 107 SerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeuAlaLeuSerLeuGlyVal 126
Db 391 GCGGATATGACGCGCATCGCTTACGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 450
Qy 127 GlyLeuSerPheSerLeuGluLeuValGlyGluLeuLysProPheIleProIleHis 146
Db 451 GGGGCTTCTGTCGCGACGAGCGGCTTCTGCGGCGGCTGCGGCGGCTGCGGCGGCTG 510
Qy 147 HisMetGluAlaHisIleAlaLeuThrIleArgLeuThrAsnLys---ValGluPheProPhe 165
Db 511 CACATCGAAGGCGCATCTGCTGCGCGGATCGCTGGAAGAGACGACCGCGCTTCCGCTTC 570
Qy 166 LeuValLeuLeuIleSerGlyIleHisCysLeuLeuAlaLeuValGluGlyValSerAsp 185
Db 571 GTCGCTTCTGCTGCTTCCGCGGCTGACACCGCATGCTGCGGCTGCGGCTGCGGCTG 630
Qy 186 PheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAla 205
Db 631 TACCAAGTTCCTGCGCAATCGGTGACGATGCGCGGCGGCAACCTTTCGACAGACCGCC 690
Qy 206 ArgAspLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysAlaIle 225
Db 691 AAGCTATCGGCTG---GGCTATCCC-----GGTGGTCCGGAATC 729
Qy 226 GlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHis 245
Db 720 GCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAT 789
Qy 246 AlaLysAsnCysAspPheSerPheThrGlyLeuGluHisValThrAspLysIleIleMet 265
Db 790 CGCGCGGCGCTGACCTTCAAGCTTCAAGCGGCTCAAGACCTTTCACCTGAAC---ACCTGG 846
Qy 266 LysLysGluLysGlnGlyIleGlyLysGlyGluIleLeuSerSerAlaAlaAspIle 285
Db 847 CAGCGTGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGATC 897
Qy 286 AlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIle 305
Db 898 GCGCTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 957
Qy 306 LeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGly 325
```

```
Db 958 -----AAGCAGACCGGCGCTG-----AAGAAC-----CTGGTGAATCGCGCG 993
Qy 326 GlyValAlaSerAsnThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGln 345
Db 994 GGTGTAAAGCGCAACAGCGCGCTGCGGCGGCTGGAAGAATGCTCGCGCAAAATGAAG 1053
Qy 346 CysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrp 365
Db 1054 GGGCAGGTGTTTCAACCGCGCGCGCGCGCTTGTGACCGACATGCGCGGATGATCGCTTAC 1113
Qy 366 AsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArg 385
Db 1114 GCGCGCTGCGGCGGCGGCTGCGCGCGG-----CAGCATGCGGCGCGCGATCAGC 1164
Qy 386 TyrGluProLysCysProLeu 392
Db 1165 GTCAGACCGCGCTGCGGATG 1185
```

## RESULT 9

```
US-09-252-991A-884/C
; Sequence 884, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 884
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-884
```

## Alignment Scores:

Pred. No.:	1,526-52	Length:	1059
Score:	492.50	Matches:	128
Percent Similarity:	53.24%	Conservative:	61
Best Local Similarity:	36.06%	Mismatches:	143
Query Match:	23.18%	Indels:	23
DB:	4	Gaps:	9

US-10-649-273-2 (1-414) x US-09-252-991A-884 (1-1059)

```
Qy 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGly 58
Db 1038 GTGCTGGAGCTGGAAGAGCTGCTGCGAAGAAACCGCGCTGCTTACGACGGAACGC 979
Qy 59 AsnValLeuGlyAlaIleHisSerGlnThrGluValHisLysLysThrGlyIle 78
Db 978 GGCCTGTGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 919
Qy 79 ValProProAlaIleGluThrLeuHisArgGluAsnIleGlnArgIleValGlnGluAla 98
Db 918 GTGCCCGAGCTGCGCTGCGCGGACCATGCAAGCGCATGCGCGGATGCGCGGAGTGG 859
Qy 99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysPro 118
Db 858 CTGACAGAGTCCGCGCTGACGCGCGGAGATATGACGCGATGCGCTTATACCGCGGCTCC 799
Qy 119 GlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeu 138
Db 798 GGCCTGTGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 739
Qy 139 LysLysProPheIleProIleHisIleMetGluAlaHisIleAlaLeuThrIleArgLeuThr 158
```





```
Db      946 ACCGATATGATGTCATGATCTTTGGCGGGATGATCCGTTTAAAGTGTACCGAG 1005
Qy      376 GYIleuHISAspIleGluGlyIleArgTyrGluProLysCysProLeu 392
      1006 GGGGCTTTA-----GGGTGACAGTGAACACGTTGGCTTTA 1044

RESULT 11
US-09-902-540-6612
; Sequence 6612, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 6612
; LENGTH: 996
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-6612

Alignment Scores:
Pred. No.:      3,01e-51      Length:      996
Score:          482.00      Matches:      128
Percent Similarity: 52.48%      Conservative: 52
Best Local Similarity: 37.32%      Mismatches: 131
Query Match:    22.68%      Indels:      32
DB:             4           Gaps:          9

US-10-649-273-2 (1-414) x US-09-902-540-6612 (1-996)
Qy      38 ILeValleuGlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThr 57
      ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      4 CTCGCTCAGAGACTGCGAAACCTCGTGTATGAGACTGCCGCCCTCGTGAGGACGCG 63
Qy      58 GYanValleuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyGly 77
      ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      64 CGCCGCGGCTTGTGATGTCGTCCTCCACGAGTGAACATCCACCGCGGTGGG 123
Qy      78 ILeValProProAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGlu 97
      ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      124 GTGTCGCGGAGCTGGCCGACCGCAACCATCTCCAGGTGCTGCCCTCGTCCACGAG 183
Qy      98 AlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLys 117
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      184 GCGGTACGCGGCGGCGAACAAGCGCTCGACGACGTGACCTCATCGCCCTCGACGTCGCG 243
Qy      118 ProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnValGlyGln 137
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      244 CCCGACTCATCGCGCGCTGCTGTGGAGTGCAGTGCAGGCGCAAGGCTTGAGCTGCGCG 303
Qy      138 LeuLysProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeu 157
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      304 ACGGCGAAGCCCTTCTGGGCGCCCAACACCTGAGAGGCGACCTGCTGCCATCCGGCTG 363
Qy      158 ThrAsnLysVal---GluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeu 176
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      364 TTGAGGTGGCGCCGAGCGCGCGCTTCTTGGGCTCGTGGTCCGCGCGGACACACAGC 423
Qy      177 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAla 196
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      424 CTCCTCAGAGTGCAGGCTTACGCGGCACTACCGGCTGTGGCGGACGACGCGACGACGCG 483
Qy      197 ProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCys 216
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      484 GCGGCGAGGATATGACAAAGACCGCTCGCATCTCGGCTG---CCGTATCCG----- 534
```

```
Qy      217 SerThrMetSerGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsn----- 233
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      535 -----GGTGGGAGCCCATCGACCAAGTTCGGCGAGCGGAAACCGGAGGCC 582
Qy      234 ---ArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSer 252
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      583 ATCCGCTTC-----CCGCGCGCTCCCGGCGCAACAATTCCAGCTGTCC 627
Qy      253 PheThrGlyLeuGlnHisValThrAspLysIleIleMetLysGluGluGluGly 272
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      628 TTCTCCGGTTTGAAG-----ACGGCGGTCTGCACCACTGCGAAGACAGCG 675
Qy      273 ILeGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThr 292
      ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      676 GTGCCCGAGCGGAGCGCGCTG-----GCCGATTGTGGCGCGTCTCCACGAGAGCC 726
Qy      293 MetAlaCysHisLeuValLysValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAsp 312
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      727 GTGCGCGAGCGTGTGTCGAAGAAG-----CTGGTGGCCGCGCGCGCGG 771
Qy      313 LeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyr 332
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      772 TTG-----GCCACAGACAGTTGTGTGTGTGCGCGCGCGCTCGCGGAATCGCGG 822
Qy      333 ILeArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProPro 352
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      823 CTCGCGGACACTGTCTACGGCGCGGCGGAGGAGCGGCGGTGAACAATCTTCGCCCGG 882
Qy      353 ProArgLeuCysThrAspAsnGlyIleMetIleAlaIleTyrAsnGlyIleGluArgLeuArg 372
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      883 GTGCGGTGTGCAACGACATATGCGCGGTATGCGGGGTATGAGCGGTACCGG 942
Qy      373 AlaGlyLeu 375
      :|||||:
Db      943 CGCGGCTCG 951

RESULT 12
US-09-902-540-503/C
; Sequence 503, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 503
; LENGTH: 2582
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-503

Alignment Scores:
Pred. No.:      1.62e-50      Length:      2582
Score:          482.00      Matches:      128
Percent Similarity: 52.48%      Conservative: 52
Best Local Similarity: 37.32%      Mismatches: 131
Query Match:    22.68%      Indels:      32
DB:             4           Gaps:          9

US-10-649-273-2 (1-414) x US-09-902-540-503 (1-2582)
Qy      38 ILeValleuGlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThr 57
      ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1525 CTCGCTCAGAGACTGCGAAACCTCGTGTATGAGACTGCCGCCCTCGTGAGAGACGCG 1466
```

```

Oy      58  GlyAsnValLeuGlnValAlaIleHisSerGlnThrGluValHisLeuValThrGly 77
Db      1455  CGCCGCGCGCTTCCGATGTCGCTCCGACGAGTGAATCAACCGCGCGGGGGGTGG 1406
Oy      78  IleValProProlAlaGlnIleuHisArgGluAsnIleGlnArgIleValGlnGlu 97
Db      1405  GTGGTGCCGAGCTGGCCGACCGGCAACACATGTCGACGATCTCCGCTGCTCCACGAG 1346
Oy      98  AlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLys 117
Db      1345  GCGCTGACGCGGCGCAACAGACGCTCGACGACGCTGATGCGCTGACGCTGACGCTCCG 1288
Oy      118  ProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGln 137
Db      1285  CCGGACCTCATCGCGCGCTGCTGTGGAGAGTCAGTGGCCGACGCGCTTGAAGCTTGGCG 1228
Oy      138  LeuLysProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeu 157
Db      1225  ACGCGCAAGCCCTTGTGGGGCCCAACACCTCGAGAGGCCACCTGCTGCCATCCGCTG 1166
Oy      158  ThrAsnLysVal--GluPheProPheLeuValLeuLeuIleSerGlyValHisCysLeu 176
Db      1165  TTGGAGTGTCGCGCCGAGCGCGCTTCCTGGGCTGTCGTTCGCGGGGACACAGC 1106
Oy      177  LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleLeu 196
Db      1105  CTGACGAGTGACGAGCCCTACCGGACGATGCGCGTGTGGGACGACGCGGACGACCGC 1046
Oy      197  ProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCys 216
Db      1045  GCCGCGACGAGCATATGACAGACCGCTGCGATCTCCGCTG--CGTATCCG----- 995
Oy      217  SerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsn----- 233
Db      994  -----GGTGGGACGCCCATGACCGAGTTGGGCGACAGGGGAAACCGGAGGCC 947
Oy      234  ---ArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSer 252
Db      946  ATCCGCTTC-----CCGCGCGCGCTGCCGCGGACACATTCGACGTGTCC 902
Oy      253  PheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGluLysGlnGluGly 272
Db      901  TTCTCCGGGTTCGAG-----ACGCGCGTCTGTCACACGTCGACAGACAGCGC 854
Oy      273  IleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThr 292
Db      853  GTGCCGCGAGGGGACAGCGCTG-----GGCATTTGTGGCGGCTCTTCAAGAGGCC 803
Oy      293  MetAlaCysHisLeuValLysArgThrHisValArgAlaIleLeuPheCysLysGlnArgAsp 312
Db      802  GTGGGAGCAGTCTCTCGAAGAG-----CTGTGGCGCGCGCGCGCGCGG 758
Oy      313  LeuLeuProGlnAsnAspAlaValLeuValAlaSerGlyValAlaIleSerAsnPheThr 332
Db      757  TTG-----GCCACAGACAGTTGGTGTGGCGCGCGCGGTGCGGCAACTCCGCG 707
Oy      333  IleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProPro 352
Db      706  CTCGCGGACGCTGTCCAGCGCGACCGACGAGAGCGGGGTTGAACATTTCTGCCCCG 647
Oy      353  ProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArg 372
Db      646  GTGCGCGCTGTGCACGCAATGCGCCCATGTGATGCGTGCGGGGTATGAGGCGTACCG 587
Oy      373  AlaGlyLeu 375
Db      586  CGCGGCTTG 578

RESULT 13
US-08-087-797-1
; Sequence 1, Application US/08087797
; Patent No. 5543312
; GENERAL INFORMATION:

```

```

1 APPLICANT: Mellors, Alan
2 APPLICANT: Lo, Reggie Y.C.
3 APPLICANT: Abdullah, Khalid M.
4 TITLE OF INVENTION: Pasteurella Haemolytica
5 TITLE OF INVENTION: Glycopolase
6 TITLE OF INVENTION: Gene and the Purified Enzyme
7 NUMBER OF SEQUENCES: 3
8 CORRESPONDENCE ADDRESS:
9 ADDRESSER: Bell, Seltzer, Park & Gibson, P.A.
10 STREET: 121 East Morehead Street,
11 CITY: Charlotte
12 STATE: No. 554312th Carolina
13 COUNTRY: United States
14 ZIP: 28234
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: Patentin Release #1.0, Version #1.25
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/087, 797
22 FILING DATE: 14-JUL-1993
23 CLASSIFICATION: 435
24 ATTORNEY/AGENT INFORMATION:
25 NAME: Layton, Jr., Samuel G.
26 REGISTRATION NUMBER: 22807
27 REFERENCE/DOCKET NUMBER: 3374-B0
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: 704 334 1561
30 TELEFAX: 704 334 2014
31 INFORMATION FOR SEQ ID NO: 1:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 1315 base pairs
34 TYPE: nucleic acid
35 STRANDEDNESS: single
36 TOPOLOGY: linear
37 MOLECULE TYPE: cDNA
38 HYPOTHEITICAL: NO
39 ANTI-SENSE: NO
40 FEATURE:
41 NAME/KEY: CDS
42 LOCATION: 141..1115
43 US-08-087-797-1
44 Alignment Scores:
45 Pred. No.: 1.19e-50 Length: 1315
46 Score: 479.00 Matches: 128
47 Percent Similarity: 48.63% Conservative: 50
48 Best Local Similarity: 34.97% Mismatches: 157
49 Query Match: 22.54% Indels: 31
50 DB: 1 Gaps: 8
51 US-10-649-273-2 (1-414) x US-08-087-797-1 (1-1315)
52 QY 9 GlyValPhepHeLyProSerIysArgLysValTyrGluPheLeuArgSerPheAsnPhe 28
53 ||| .||||||| ||| :||| ||| |
54 89 GGGGCGCTTCGCCCTC-----TTGGTTTTCAACTATATTGTGACTTC 133
55 QY 29 HisPrOgLYThrLeuPheunHisLysIle-VallengUYllGLUThrSeryGapas 48
56 -----AACTATGCgaatttTAGGaTtaAGaaCCttTGTAATG:::
57 134 TCC-----ACCTATGCgaatttTAGGaTtaAGaaCCttTGTAATG 175
58 QY 48 pThraAlaAlaValaIsPGlUthrGYlaenVallengUYgluaalIethIsserGI 68
59 ||| |||:| ||| | :||| ||| |
60 176 AAACCGctgttcCATTTATGAtGAAACAAGCGTTAgTgcCAAACAGCTTATAGCCA 235
61 QY 68 nThrgIuVaHlsleuLySThrGlyGlYLleValPRoPRoaLaIlagIngnIleuhISar 88
62 ::::::||| ||| ||| ||| ||| ||| ||| |||
63 236 AATmATATgCaGCGCGAAttAGGtgGAGCGTAAGccCTGAaAcTcgGCTTCGAGAcaTAT 259
64 Db 88 gGUaaenlleGnaaglileValInglnGuLaaleuserrallaSerdlvalserProSerAs 1088
65 ::::::||| ||| ||| ||| ||| ||| ||| |||
66 296 CGGTAAAAGTTGCCCATTAATTCAAGAACCTTTAAAGAGGCGCAAnTCyGmaCCCTCGGA 3555

```



Db 552728 GCGAAGCCTTGATTAACAGCAAAATTACTTGACTA---GATTATCCA----- 552681  
Qy 218 ThrmSerGlyGlyValAlaIleGluHisLeuAlaIleGlnIValAsnArgPheHisPhe 237  
Db 552680 -----GGTGGCGCGGACCTTCTCGTTTACGGGAAAAAGGTAAGCCAAATCGTTTC 552630  
Qy 238 AspIleValProPheLeuHisHisAlaIleValAsnGlyAsnGlySerPheThrGlyLeuGln 257  
Db 552629 ACATTTCACGTCGATGACAGATCGTCGAGGCCCTTGATTTTGTGTTTAA 552570  
Qy 258 HisValThrAspIleIleMetLeuValGluGluGluGly-----IleGluVal 275  
Db 552569 ACATTGCCCGCAATACATTAATCAAGCAATTAAAAACGAGGCGCAATGATGAGAGCA 552510  
Qy 276 GlyGlnIleLeuSerSerAlaIleAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 295  
Db 552509 -----ACTAAGAGATATTGCTTATGCTTTCGAAAGATGGCGGTGGAT 552465  
Qy 296 HisLeuValIleValArgThrHisValGlnAlaIleLeuPheCysLeuGlnArgAspLeuPro 315  
Db 552464 ACTCTTGCC-----ATTAATGTAG---CGTCATTGAAAGAA 552429  
Qy 316 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheThrIleArgArg 335  
Db 552428 ACAGGCTATAACGTTTATGATTCGGGAGGCGTACGCAATAAAACTCCGAGAA 552369  
Qy 336 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 355  
Db 552368 ACGCTTGCGCATTAATGCAAAATTAGTGGCGAAGTGTATTATCTCAACCTCAATTT 552309  
Qy 356 CysThrAspAsnGlyIleMetIleAlaIleAlaIleAlaIleGluArgLeuArgIleGly 374  
Db 552308 TGTACAGATTAATGTCGATGATGCTTACACAGGTTTTTTCAGTTTAAACAAAGGT 552252

RESULT 15  
US-09-643-990A-1/c  
Sequence 1, Application US/09643990A  
Patent No. 6528289  
GENERAL INFORMATION:  
APPLICANT: Robert D. Fleischmann  
Mark D. Adams  
Owen White  
Hamilton O. Smith  
J. Craig Venter  
TITLE OF INVENTION: The Nucleotide sequence of  
Thereof, and Uses Thereof  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville,  
STATE: MD  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS v6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/643,990A  
FILING DATE: 23-Aug-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/487,429  
FILING DATE: 1995-06-07  
APPLICATION NUMBER: 08/426,787  
FILING DATE: 1995-04-21  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB186PIC1

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-610-5790  
TELEFAX: 310-309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-643-990A-1  
Alignment Scores:  
Pred. No.: 1,21e-44 Length: 1830121  
Score: 475.50 Matches: 120  
Percent Similarity: 50.74% Conservative: 52  
Best Local Similarity: 35.40% Mismatches: 144  
Query Match: 22.38% Indels: 23  
DB: 4 Gaps: 7  
US-10-649-273-2 (1-414) x US-09-643-990A-1 (1-1830121)  
Qy 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGly 58  
Db 553208 ATCTTAGCATTGTAATTAACCTTCTGATGAAACGGCGCTGGCATTTAGTGAAGAAAA 553149  
Qy 59 AsnValLeuGlyGlnAlaIleHisSerGlnThrGluValHisLeuValSerThrGlyIle 78  
Db 553148 GGAATTAATTCCTAATCACTTATATCTCAAAATTCCTCGCATGCAATTAATGCTGTGTG 553089  
Qy 79 ValProProAlaIleGlnIleLeuHisArgGluAsnIleGlnArgIleValGlnGlnAla 98  
Db 553088 GTTCTGTAATTAGATGATCAGCGATCATATTCGAAAACAGACCTCTTATTAACGGCA 553029  
Qy 99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleValPro 118  
Db 553028 TTAGAGAGAACCAATTAAACCGCAGCATATGATGATGCTTATACGATGGCCCT 552969  
Qy 119 GlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeu 138  
Db 552968 GGGCTGTGGCGCATGCTGTGCGTGCATGTCACATTCGACGTTCTTAAGCTATGCTTG 552909  
Qy 139 LysLeuProPheIleProIleHisIleMetGluAlaHisAlaIleThrIleArgLeu--- 157  
Db 552908 AAGTCTCTCGCAATGCTGTTCATATGAAAGGCAATTACTTGCGCCAAAGCTTGAT 552849  
Qy 158 ThrAsnValValGluPheProPheLeuValLeuIleSerGlyGlnHisCysLeuLeu 177  
Db 552848 GACAAATTCACCGCACCTTCTTTGTGCTGTGGTATCGGGTGCACACATCAATTA 552789  
Qy 178 AlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaPro 197  
Db 552788 GTCGGTGTCATGATGTAGGAAATATGAATGATAGCAAACTATGATGATGCTGCT 552729  
Qy 198 GlyAspMetLeuAspIleValAlaArgLeuSerLeuIleLysHisProGluCysSer 217  
Db 552728 GCGAAGCCTTGATTAACAGCAAAATTACTTGACTA---GATTATCCA----- 552681  
Qy 218 ThrmSerGlyGlyValAlaIleGluHisLeuAlaIleGlnIValAsnArgPheHisPhe 237  
Db 552680 -----GGTGGCGCGGACCTTCTCGTTTACGGGAAAAAGGTAAGCCAAATCGTTTC 552630  
Qy 238 AspIleValProPheLeuHisHisAlaIleValAsnGlyAsnGlySerPheThrGlyLeuGln 257  
Db 552629 ACATTTCACGTCGATGACAGATCGTCGAGGCCCTTGATTTTGTGTTTAA 552570  
Qy 258 HisValThrAspIleIleMetLeuValGluGluGluGly-----IleGluVal 275  
Db 552569 ACATTGCCCGCAATACATTAATCAAGCAATTAAAAACGAGGCGCAATGATGAGAGCA 552510  
Qy 276 GlyGlnIleLeuSerSerAlaIleAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 295  
Db 552509 -----ACTAAGAGATATTGCTTATGCTTTCGAAAGATGGCGGTGGAT 552465

```

Qy      296 HisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuPro 315
Db      552464 ACTCTTGCC-----ATTAATGTAAG---CGTCATTGAAAGAA 552429
Qy      316 GlnAsnAsnAlaValIleuValAlaSerGlyValAlaSerAsnPheTyrIleArgArg 335
Db      552428 ACAGGCTATAAAGCTTGTAGTGATGCGGAGGGGTGACCGCAATATAAAACTCCGAGAA 552369
Qy      336 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 355
Db      552368 ACGCTTGCGCACTTAATGCATAATTAGGTGGCGAAGTGTATTATCTCAACCTCAATTT 552309
Qy      356 CysThrAspAsnGlyIleuTyrIleAlaTyrAsnGlyIleGluArgLeuArgAlaGly 374
Db      552308 TGTACAGATTAATGTGTGCGATGATTGCTTACACAGGTTTATTAGTTTAAACAAAGGT 552252

```

Search completed: November 11, 2005, 02:47:48  
 Job time : 1671.85 secs





```

Db 471 CCAGAGCTCAACAGCTTCAAGAGAAATATTCAGATAGTACAGAAAGCTCTTCT 530
Qy 101 AlAserGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLeu 120
Db 531 GCCAGTGGAGTCTCCAGAGTACCTTCACAGCAATGACATACATAAACAGAGACTT 590
Qy 121 AlAeuSerLeuGlyValGlyLeuSerPheSerLeuGlyLeuValGlyLeuLeuLeu 140
Db 591 GCTTAAAGCTGGAGTGGAGTCTTATTCATTACCTTACCTGGTGGTGGACATTAAG 650
Qy 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeu 160
Db 651 CCATTCAATTCATTCAATTCATATGAGAGCTCATGACCTACTATAGTTAGTGCATATA 710
Qy 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisGlyLeuLeuAlaLeuVal 180
Db 711 GTAGAAATTTCTTTTGTAGTTCTTTGTAGTTCTGAGAGTCACTGCTGTTGGCATTAAGT 770
Qy 181 GluGlyValSerAspPheLeuLeuLeuGlyLeuSerLeuAspIleAlaProGlyAspMet 200
Db 771 CAAGAGATTTCAAGATTTCTGCTTGTGAAAGTCTTGGACATAGCACAGAGTGCATG 830
Qy 201 LeuAspLeuValAlaArgArgLeuSerLeuIleLeuHisProGlyCysSerThrMetSer 220
Db 831 CTTGACAGGTGGCAGAGAGCTTTCTTAATTAACATCCAGAGTGCCTCCACATGAGT 890
Qy 221 GlyGlyLeuValIleGlyHisLeuAlaLeuGlyGlyAsnArgPheHisPheAspIleLeu 240
Db 891 GGTGGAGAAAGCCATGAGACATTTGGCCAAACAGAAATAGATTCAATTTGACATCAAA 950
Qy 241 ProProLeuHisHisAlaLeuAsnGlyAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 951 CCTCCCTTGCAATCAAGTAAATGTAATTTCTTTTACCTGAGCTTCAACACGTTACT 1010
Qy 261 AspLeuIleIleMetLeuLeuGlyLeuGlyGlyGlyIleGlyLeuGlyGlyIleLeuSer 280
Db 1011 GATTAATAATATATATAAAGAGAAAGAGAGATATGAGAGAGGCGAAATCTGTCT 1070
Qy 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLeuArg 300
Db 1071 TCAGAGAGAGATGCTGCTGCCACATGACACACATGAGCATGCTATTTGAGAAAG 1130
Qy 301 ThrHisArgAlaIleLeuPheCysLeuGlnArgAspLeuLeuProGlnAsnAlaVal 320
Db 1131 ACACATCGGGCTATCTGTTTGTAAAGCAGAGAGCTTGTATCTTCAATAATATGACATA 1190
Qy 321 LeuValAlaSerGlyValAlaSerAsnPheThrIleArgArgAlaLeuGluIleLeu 340
Db 1191 CTGGTGCATCTGGGGTGTGCGAAGTAACTTATATCCGAGAGCTCTGGAAATTTTA 1250
Qy 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
Db 1251 ACAACGCAACACAGTGCATTTGTTGTCTCTCCCTCCAGACTAGCATGATAATGG 1310
Qy 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db 1311 ATTATGATTCAGTGAATGATGTAATGAAAGACTACGCTGCGCTGGCGCATTTTACATGAC 1370
Qy 381 IleGlyGlyIleArgGlyGluProLeuCysProLeuGlyValAspIleSerLeuGlyVal 400
Db 1371 ATAGAGAGCATCGCTATGAAACCAAAATATCTCTTGGAGTACATATCAAAAGAGATT 1430
Qy 401 GlyGlyAlaSerIleLeuValProGlnLeuLeuMetGluIle 414
Db 1431 GGAGAGCTTCCATTAAGTACCAATTAATAATGAGATA 1472

```

## RESULT 2

```

US-10-649-273-1
; Sequence 1, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1

```

```

; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231) .. (1472)
US-10-649-273-1
Alignment Scores:
Pred. No.: 0 Length: 2197
Score: 414.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 19
US-10-649-273-2 (1-414) x US-10-649-273-1 (1-2197)
Qy 1 MetLeuIleLeuThrIleThrAlaGlyValPhePheLeuProSerLeuArgLeuValTyr 20
Db 231 ATGCTAATCTTGAATCAATGAGTGCAGAGATTTTATTAACCAACAAAGAAAGTTAT 290
Qy 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40
Db 291 GAATTTTAAAGATTTTAAATTTTCACTCGAAGCACTATTTCTTCAATAAATAGTATG 350
Qy 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaLeuGluThrArgIleAsnVal 60
Db 351 GGAAATGAAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 410
Qy 61 LeuGlyGluAlaIleHisSerGluThrGluValHisIleLeuThrGlyGlyIleValPro 80
Db 411 TTGGAGAGCAATATCAATTCCTCAAACTGAACTTATTAATAACAGGTGGATTTCTCT 470
Qy 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Db 471 CCAGAGCTCAACAGCTTCAAGAGAAATATTCAGAGATAGTACAGAAAGCTCTTCT 530
Qy 101 AlAserGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLeu 120
Db 531 GCCAGTGGAGTCTCTCCAGAGTCACTCTCAGCAATGGCACTCCATTAACCAAGAGACTT 590
Qy 121 AlAeuSerLeuGlyValGlyLeuSerPheSerLeuGlyLeuValGlyLeuLeuLeu 140
Db 591 GCTTAAAGCTGGAGTGGAGTCTTATTCATTACCTTACCTGGTGGTGGACATTAAG 650
Qy 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeu 160
Db 651 CCATTCAATTCATTCAATTCATATGAGAGCTCATGCACTTATCTATTAAGTGCATATA 710
Qy 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisGlyLeuLeuAlaLeuVal 180
Db 711 GTAGAAATTTCTTTTGTAGTTCTTTGTAGTTCTGAGAGTCACTGCTGTTGGCATTAAGT 770
Qy 181 GluGlyValSerAspPheLeuLeuLeuGlyLeuSerLeuAspIleAlaProGlyAspMet 200
Db 771 CAAGAGATTTCAAGATTTCTGCTTGTGAAAGTCTTGGACATAGCACAGAGTGCATG 830
Qy 201 LeuAspLeuValAlaArgArgLeuSerLeuIleLeuHisProGlyCysSerThrMetSer 220
Db 831 CTTGACAGGTGGCAGAGAGCTTTCTTAATTAACATCCAGAGTGCCTCCACATGAGT 890

```

QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
DB 891 GGTGGAAAGCCATGAAACATTGGCCAAACAGAAATAGATTTCATTTCATTCACATCAAA 950  
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB 951 CTTCCCTTGCATCAATGCTTAAATATGTTGATTTTCTTTTACGTGACTTCAACACGTTACT 1010  
QY 261 AspLysIleIleMetLysLysGlnLysGlnGlnGlyIleGlnLysGlyGlnIleLeuSer 280  
DB 1011 GATAAATAATATATGAAAAAGAAAAAGAGATATTGAGAAGGGCAAAATCTGTCT 1070  
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
DB 1071 TCACACAGACAGATTCGTCCACAGTACACACAAATGCAATGATCATCTTGTGAAGAAGA 1130  
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320  
DB 1131 ACACATCGGGCTATTTCTGTTTGTGAAGACAGAGACTGTGTATCTCAAAATATATCAGTA 1190  
QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeu 340  
DB 1191 CTGGTTGCATCTGGTGTGTGTCCAGTAACTTCTATATCCGACAGCTCTGGAATTTTA 1250  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360  
DB 1251 ACAACGCAACACAGTCACATTGTTGTCTCTCCACAGCTATGACATATATATGAC 1310  
QY 361 IleMetIleAlaTyrAsnGlyIleGlnArgLeuAlaGlyLeuGlyIleLeuHisAsp 380  
DB 1311 ATATATGATGATGATGAAATGTAATGAAAGACTACCTGCTGCGCTTGGCATTTTACATGAC 1370  
QY 381 IleGlnGlyIleArgTyrGlnProLysCysProLeuGlnValAspIleSerLysGlnVal 400  
DB 1371 ATAAAGGCATCCGCTATGAACCAAAATGTCCTCTTGAAGTACATATCAAAAGAGTT 1430  
QY 401 GlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 414  
DB 1431 GGAGAGCTTCATTAATAAGTACCAATTAATAAGAGATA 1472

RESULT 3  
US-10-651-722-1  
; Sequence 1, Application US/10651722  
; Publication No. US20040048302A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 DIV  
; CURRENT APPLICATION NUMBER: US/10/651,722  
; PRIOR FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1  
; LENGTH: 2197  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (231)..(1472)  
US-10-651-722-1

Alignment Scores:  
Preg. No.: 0 Length: 2197  
Score: 414.00 Matches: 414  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0  
DB: 19 Gaps: 0  
US-10-649-273-2 (1-414) x US-10-651-722-1 (1-2197)

QY 1 MetLeuIleLeuThrIleThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
DB 231 ATGCTATCTTGTGCTAAGCTGACAGAGATTTTTTTTTTAAACATCAAAAGAAAGATTAT 290  
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
DB 291 GAATTTTAAAGATTATTAATTTTCAATCCGGAACATAATTTCTTCAATAAATAGTAATG 350  
QY 41 GlyIleGlnThrSerCysAspAspThrAlaAlaValValAspGlnThrGlyAsnVal 60  
DB 351 GGAATTAAGAACTAGTTGTGTATGATACAGACGCTGCTGTGTGATGAATCTGGAATATG 410  
QY 61 LeuGlyGlnAlaIleHisSerGlnThrGlnValHisLeuLysThrGlyIleValPro 80  
DB 411 TTGGGAGAAAGCAATATACATTTCCAACTGAAGTCAATTAATAAAGAGGTGGATTGCTCT 470  
QY 81 ProAlaAlaGlnGlnLeuHisArgGlnAsnIleGlnArgIleValGlnGlnAlaLeuSer 100  
DB 471 CCAAGCTCAACAGCTTCAACAGAAATATTCACGAATATGACAAAGAGCTCTTTCT 530  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
DB 531 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATTCGCAATACATTAACCAAGAGACT 590  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyLeuLysLys 140  
DB 591 GCTTTAAGCTGGAGTGGCTTATCATTTAGCTTACAGCTGTATGAGACGTTAAAAAG 650  
QY 141 ProPheIleProIleHisMetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
DB 651 CCAATTCATTCATTCATATGATGAGGCTCAAGCACTTATCTTTAGTTGACCAATAA 710  
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
DB 711 GTAGATTTCTTTTATTTAGTTCTTTGATTTCTGGAGGTCACTGTCTGTTGGCATTAAGT 770  
QY 181 GlnGlyValSerAspPheLeuLeuGlnLysSerLeuAspIleAlaProGlyAspMet 200  
DB 771 CAAGAGTTTCAAGATTTTCTGCTTCTTGAAGTCTTTGGAATACACACAGGTGACATG 830  
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlnCysSerThrMetSer 220  
DB 831 CTGACAAAGTGGCAAGAACTTTCTTAAATTAACATCCAGAGTCTCCACATGAGT 890  
QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
DB 891 GGTGGAAAGCCATGAAACATTGGCCAAACAGAAATAGATTTCATTTCATTCACATCAAA 950  
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB 951 CTTCCCTTGCATCAATGCTTAAATATGTTGATTTTCTTTTACTGACTTCAACACGTTACT 1010  
QY 261 AspLysIleIleMetLysLysGlnLysGlnGlnGlyIleGlnLysGlyGlnIleLeuSer 280  
DB 1011 GATAAATAATATATGAAAAAGAAAAAGAGATATTGAGAAGGGCAAAATCTGTCT 1070  
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
DB 1071 TCACACAGACAGATTCGTCCACAGTACACACAAATGCAATGATCATCTTGTGAAGAAGA 1130  
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320  
DB 1131 ACACATCGGGCTATTTCTGTTTGTGAAGACAGACTGTGTATCTCAAAATATATCAGTA 1190  
QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeu 340  
DB 1191 CTGGTTGCATCTGGTGTGTGTCCAGTAACTTCTATATCCGACAGCTCTGGAATTTTA 1250

```
OY 341 ThrAsnaIaThrGlnCysThrIeuCysProProArgLeuCysThrAspaSngly 360
Db 1251 ACAAAACGACACAGTGCATCTTGTGTCTCTCCAGACTATGACATATATGGC 1310
OY 361 ILeuEIIealIaTPaSnglyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHIsaP 380
Db 1311 ATTATGATTGATGTAATGGTATGAAGACTACGCTGCTGGGCAATTTTACATGAC 1370
OY 381 ILeuGlyIleArgTyrgLuprolyscysProleuGlyValAspIleSergIVal 400
Db 1371 ATAGAAAGCATCCGCTATGAACCAAAATCTCTTGGAGTACATATCAAAAGAGTT 1430
OY 401 GlyGluAlaSerIleLysValaProGlnLeuLysMetGluIle 414
Db 1431 GGAGAGCTTCATAAAGTACCACAAATTTAAATGAGATA 1472

RESULT 4
US-10-480-988-36
; Sequence 36, Application US/10480988
; Publication No. US20050069877A1
; GENERAL INFORMATION:
; APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;
; APPLICANT: SWARNAKAR, Anita; HAFALIA, April J.A.;
; APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;
; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;
; APPLICANT: HONCHELL, Cynthia D.; NGUYEN, Daniel B.;
; APPLICANT: LU, Dyanh Aina M.; LEE, Ernestine A.;
; APPLICANT: YUE, Henry; FORSYTHE, Ian J.;
; APPLICANT: BARROSO, Ines; RAMKUMAR, Jayalaxmi;
; APPLICANT: GRIFFIN, Jennifer A.; Li, Joana X.;
; APPLICANT: YANG, Junming; THANGAVELU, Kavitha;
; APPLICANT: GIETZEN, Kimberly J.; DING, Li;
; APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.;
; APPLICANT: YAO, Monique G.; CHAWLA, Nandinder K.;
; APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
; APPLICANT: LEE, Sally; BECHA, Shanya D.;
; APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;
; APPLICANT: ELIOTT, Vicki S.; LEO, Wen;
; APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;
; APPLICANT: LU, Yan; ZEBARJADIAN, Yeganeh
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PF-1040 USN
; CURRENT APPLICATION NUMBER: US/10/480,988
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: PCT/US02/19360
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/300,508
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/303,445
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/305,405
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/311,442
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 60/314,821
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/315,992
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 60/378,205
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 36
; LENGTH: 2572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 7632424CB1
US-10-480-988-36

Alignment Scores: 0 Length: 2572
Pred. No.: 3
```

```
Score: 414.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0
US-10-649-273-2 (1-414) x US-10-480-988-36 (1-2572)

OY 1 MetLeuIleLeuThrIleThrAlaGlyValaPhePheLysProSerysArgLysValaTyr 20
Db 144 ATGCTAATTTTGACTAAGACTGAGAGATTTTTTTTAAACCATCAAAAAGAAATTTAT 203
OY 21 GluPheLeuArgSerPheAsnPhenIserProGlyThrLeuPheLeuHIsyIleVal 40
Db 204 GAATTTTAAAGAGTTTAAATTTTCACTCGAAGACTATTTCTCATAAATATGATTTG 263
OY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValaValaAspGluThrGlyAsnVal 60
Db 264 GGAAATTGAACCTAGTTGTGATGATACAGACGCTGCTGTGATGAAACTGAAATGTG 323
OY 61 LeuGlyGluAlaIleHisSergInThrGluValaHIsLeuLysThrGlyGlyIleValPro 80
Db 324 TTGGAGAAAGCAATACATTTCCAACTGAAGTTTATTAACACAGGTGGATTTCTCT 383
OY 81 ProAlaAlaGlnGlnLeuHIsArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Db 384 CACGACCTCAACAGCTTCCACAGAGAAATATTCAGATATATACAGAAAGCTTTTCT 443
OY 101 AlaSergIValaSerProSerysAspLeuSeraIleAlaThrThrIleLysProGlyLeu 120
Db 444 GCAGTGAAGTCTCTCAAGTACCTCAGCAATTCACATCAATCAAAACAGGACTT 503
OY 121 AlaLeuSerLeuGlyValaGlyLeuSerpheSerLeuGlyValaGlyGlnLeuLysLys 140
Db 504 GCTTTAAGCTGGAGTGGCTTATCATTTAGCTTACAGTGTAGACAGTTTAAAG 563
OY 141 ProPheIleProIleHIsIleMetGluAlaHIsAlaLeuThrIleArgLeuThrAsnLys 160
Db 564 CATTATTCCTCATTAATATGAGAGCTTCATCTATTTAGTTGACCAATAA 623
OY 161 ValGluPheProPheLeuValLeuLeuIleSergIleGlyHIsCysLeuLeuAlaLeuVal 180
Db 624 GTAGATTTCTTTTATGATTTCTTTGATTTCTGAGGTACAGTCTGTGGCATTTAGT 683
OY 181 GlnGlyValaSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
Db 684 CAAGGAGTTTCAGATTTTCTGCTTCTTGGAAAGTCTTGGACATAGCACAGGTGACATG 743
OY 201 LeuAspLysValaAlaArgArgLeuSeryIleLysIleSergIleProGluCysSeryThrMetSer 220
Db 744 CTGACAAAGGTGCAGAAAGACTTTCTTTATTAACATCCAGAGTCTCCACCATAGT 803
OY 221 GlyIleLysAlaIleGluHIsLeuAlaLysGlnIleAsnArgPheHIsPheAspIleLys 240
Db 804 GGTGGAAAGCCATAGACATTTGGCCAAACAGAAATGATTTCAATTTTGACATCAAA 863
OY 241 ProProLeuHIsHIsAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHIsValaThr 260
Db 864 CTTCCCTTGATCATGCTTAAATAATGGATTTTCTTTTACTGACATTCACACAGTTACT 923
OY 261 AspLysIleIleMetLysLysGlyLysGluGlnGlyIleGluLysGlyGlnIleLeuSer 280
Db 924 GATTAATAATTAATGAAAAAGAAAAAGAAAGGTATTATAGAAAGGGCAAAATCTGTCT 983
OY 281 SerAlaAlaAspIleAlaAlaThrValaGlnHIsThrMetAlaCysHIsLeuValaLysArg 300
Db 984 TCAGCAGACACATGCTGTCACAGTACACACACATGSCATGCTCTTGTGAAGAAGA 1043
OY 301 ThrHIsArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320
Db 1044 ACACATGGGCTATTTGTTTGTATAGCAGAGACTTGTATCCTCAAAATTAAGCAGTA 1103
OY 321 LeuValAlaSerGlyValaAlaSerAsnPhenTyrlleArgArgAlaLeuGluIleLeu 340
```

Db 1104 CTGGTTCATCTGGTGGTTCGCAAGTAATCTTATATCCGAGAGCTCTGCAAAATTTA 1163  
Qy 341 Thrasna1athrg1ncysrthrlleuencysproproproargleucysrthraspaengly 360  
Db 1164 ACAAGGCAACAGACAGTCTTGTGTCTCCCTCCAGACTAGCACTGATTAATGGC 1223  
Qy 361 llemcttiala1trpansnglyllegluargleuargalglyleuglyl1eleuhi8asp 380  
Db 1224 ATTATGATTCAGTGAATGTGTAATTGAAAGACTACGCTGGCTTGGCAATTTTACATGAC 1283  
Qy 381 l1legluglyl1eargtyrgluprolyscysproleuGlyValasp11eserlysgluval 400  
Db 1284 ATAGAGGCAATCCGCTATGAAACCAAAATGTCTTGGAGTAGACATATCAAAAGAAATT 1343  
Qy 401 Glylu1aser1lelysva1progl1neulysmetglu1le 414  
Db 1344 GGAGAAAGCTTCATTAAGTACCAATTAATAATGAGATA 1385  
RESULT 5  
US-10-120-988-177  
: Sequence 177, Application US/10120988  
: Publication No. US20030219745A1  
: GENERAL INFORMATION:  
: APPLICANT: Tang, Y. Tom  
: APPLICANT: Goodrich, Ryle  
: APPLICANT: Liu, Chenghua  
: APPLICANT: Ren, Feiyun  
: APPLICANT: Wang, Dunrui  
: APPLICANT: Drmanac, Radoje T.  
: TITLE OF INVENTION: No. US20030219745A1 Nucleic Acids and  
: FILE REFERENCE: 802CON  
: CURRENT APPLICATION NUMBER: US/10/120,988  
: CURRENT FILING DATE: 2002-04-11  
: PRIOR APPLICATION NUMBER: 09/774,528  
: NUMBER OF SEQ ID NOS: 441  
: SOFTWARE: pf FL\_genes Version 2.0  
: SEQ ID NO 177  
: LENGTH: 1416  
: TYPE: DNA  
: ORGANISM: Homo sapiens  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: (1205)..(1305)  
US-10-120-988-177  
Alignment Scores:  
Pred. No.: 0 Length: 1416  
Score: 340.00 Matches: 340  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 82.13% Indels: 0  
DB: 18 Gaps: 0  
US-10-649-273-2 (1-414) x US-10-120-988-177 (1-1416)  
Qy 75 Thr1g1gyl1leval1propro1a1a1ag1ng1neuh1sarg1l1uasn1leg1narg1le 94  
Db 283 ACAAGTGGATTTGTTCTTCACAGACTCAACAGCTTCAAGAAATATTCAACGAATA 342  
Qy 95 Valg1ng1ua1aleuSer1aseg1yValserProserAspleuSer1a1leat1thr 114  
Db 343 GTACAAAGAGCTCTTCTGCCAGTGAAGTCTCTCCAAAGTACCTTCAGCAATTCACACT 402  
Qy 115 Thr1lelyProgl1yleu1aleuSer1eug1yVal1g1yleuSerPheSer1eug1n1leu 134  
Db 403 ACCATTAACCGAGCTTGCTTAAGCTCGGAGTGGGCTTATCATTTAGCTTAACGCTG 462  
Qy 135 Valg1g1n1leuly1sProPhe1lePro1le1sh1emec1y1ua1ah1sa1leuthr 154  
Db 463 GTAGACAGTTAATAAGCAATTCATTCATTCATATGAGGCTCATGCACTTACT 522

Qy 155 l1eargleuthrasnlysva1glupheProPhe1leuVal1leuenu1eserg1y1his 174  
Db 523 ATTAGGTTGACCAATAAAGTAGAAATTTCTTTTATAGTTCTTTGATTTCTGGAGGTAC 582  
Qy 175 Cys1euleu1aleuVal1g1ng1yValSerAspPhe1leu1eug1ylysvser1eulasp 194  
Db 583 TGTCTGTGGCAATTAAGTCAAGAGATTTCAGATTTTCTTCTTGGAAAGTCTTTGGAC 642  
Qy 195 l1eal1progl1yasphe1leuaplysva1a1sarg1rgleuSer1lelysh1sPro 214  
Db 643 ATAGCACAGGTACATGCTTGACAGAGTGGCAAGAAAGCACTTTTAATAAATCAATCA 702  
Qy 215 Glu1CysSerThmetserg1ylysva1a1leg1uhi1sleu1alysg1ng1yasnarg 234  
Db 703 GAGTGTCCCAACATGAGTGTGGAAAGCCATGAACATTTGGCCAAAGAAATAGA 762  
Qy 235 Phehi1sPheasp11elysProProleuhi1sh1a1alyabancysabpPheSerPheThr 254  
Db 763 TTTCAATTTGACATCAAACTCCCTGCAATCATGCTMAAAATTTGTGATTTTCTTTACT 822  
Qy 255 Gl1y1eug1nhi1sval1thrasplys1le1lemerly1sleg1lup1s6glug1y1leg1u 274  
Db 823 GGACTTCACACGTTACTGATTAATAATATATGAAAAAGAAAGAGAAAGGATTTAG 882  
Qy 275 lyeG1yG1n1leu1ser1a1a1asp1leal1a1a1thrVal1g1nhi1sthrmet1a 294  
Db 883 AAGGGCAAAATCTGTCTTACAGCAGCAGCATTTGCTCCACAGTACAGCACATGTGCA 942  
Qy 295 Cys1e1sleuVal1ysarg1thr1sarg1a1ile1eupheCyslysg1narqasp1euleu 314  
Db 943 TGTCACTTGTGAAAAAGAACACATCGGGCTAATCTGTTTGTAAAGCAGAGACTGTGTA 1002  
Qy 315 Progl1asasna1a1a1leuVal1aseg1y1yVal1aser1asn1p1e1earg 334  
Db 1003 CTCCAAAATTAATCAGTACTGTTGCAATCTGGTGTGTCGAAGTAACTTTATATCCGC 1062  
Qy 335 Arg1aleuG1u1leu1thrasna1athrg1ncysrthrlleuencysproproproarg 354  
Db 1063 AGAGCTCTGGAATTTTAACAAACGCAACAGTGCACCTTGTGTCTCTCCACAGA 1122  
Qy 355 l1eucys1thraspaenglyl1emcttiala1trpansnglyllegluargleuargalgly 374  
Db 1123 CTATGACATGATATGATGATTAATGATTCATGGAATGATTAAGACTACGTCGGC 1182  
Qy 375 l1euglyl1leuuh1sasp11leg1uglyl1eargtyrgluprolyscysproleuGlyVal 394  
Db 1183 TTGGCAATTTTACATGACATAGAGGCAATCCGCTATGAAACCAAAATGTCTCTTGGAGTA 1242  
Qy 395 Asp11eSer1ysg1yVal1g1y1ua1aser1lelysva1Progl1neulysmetglu1le 414  
Db 1243 GACATATCAAAAGAGTTGAGAGGCTTCATTAAGTACCAATTAATAATGAGATA 1302

Search completed: November 11, 2005, 07:42:43  
Job time : 1294 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 17:34:34 ; Search time 120 Seconds

(without alignment)  
1334.323 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 414  
Sequence: 1 MLILTKTAGVFPKSKRRKY.....DISKEVGASIKVQLKMEI 414

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 2105692 seqs, 386760381 residues

Word size: 274

Total number of hits satisfying chosen parameters: 3

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: A\_Geneseq\_16Dec04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1980s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	414	100.0	414	5 ABG96478	Abg96478 Novel hum
2	414	100.0	414	5 ABB05481	Abb05481 Human O-b
3	414	100.0	414	6 ABJ26654	Abj26654 Human pro

ALIGNMENTS

RESULT 1

ABG96478 standard; protein; 414 AA.

XX ABG96478;

XX 11-DEC-2002 (first entry)

XX Novel human metalloprotease MPI.

XX Metalloprotease: MP-1; immune disorder; glutamate transport; Cancer;  
motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
liver disease; renal disease; immune disorder; rheumatoid arthritis;

KM acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
XX neurological disorder.

OS Homo sapiens.

PN MO200272751-A2.

XX 19-SEP-2002.

PF 05-FEB-2002; 2002MO-US003353.

XX 05-FEB-2001; 2001US-0266518P.

XX 10-APR-2001; 2001US-0282814P.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;

XX MPI; 2002-723329/78.

XX N-PSDB; ABS76635.

PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
PT treating, or ameliorating diseases associated with aberrant  
PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and  
PT neurological disorders.

PS Claim 5; Fig 1A-C; 473bp; English.

XX The invention describes an isolated nucleic acid molecule (I) encoding a  
CC metalloprotease (MP-1). (I) is useful for preventing, treating, or  
CC ameliorating a medical condition, particularly an immune disorder, an  
CC aberrant glutamate transport or motor neuron disorder, such as  
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
CC condition. The compositions and methods are also useful for diagnosing,  
CC prognosticating, treating, ameliorating and/or treating disorders  
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
CC sequence of a metalloprotease MPI protein

XX SQ Sequence 414 AA;

Query Match 100.0%; Score 414; DB 5; Length 414;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Gaps 0;

Matches 414; Conservative 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFPKSKRRKYEPFLRSFNHFGTLFLHKIVLGIEISCDTAAAVDRTGNV 60

DB 1 MLILTKTAGVFPKSKRRKYEPFLRSFNHFGTLFLHKIVLGIEISCDTAAAVDRTGNV 60

QY 61 LGEAHSQTEVHLKTGCIIVPPAAQOLAHRENIQRIYVDALSASGVSPDSALITTKPGL 120

DB 61 LGEAHSQTEVHLKTGCIIVPPAAQOLAHRENIQRIYVDALSASGVSPDSALITTKPGL 120

QY 121 ALSIGVGLSFSLOLVQQLKPPPIPHHMEAHALTITLNNVPEPFLVLLISGHCILAVY 180

DB 121 ALSIGVGLSFSLOLVQQLKPPPIPHHMEAHALTITLNNVPEPFLVLLISGHCILAVY 180

QY 181 QGVSPDLGKSLDIPGMDLVKARLSLIKPECSYSGGAIIEHLAKQGRFPFDIK 240

DB 181 QGVSPDLGKSLDIPGMDLVKARLSLIKPECSYSGGAIIEHLAKQGRFPFDIK 240

QY 241 PPLHAHAKNCDPFTYGLQHTYTDKIIMKEKEBEGIEKQQLISSADIAATVQHTMACHLVKR 300

DB 241 PPLHAHAKNCDPFTYGLQHTYTDKIIMKEKEBEGIEKQQLISSADIAATVQHTMACHLVKR 300

181 QGVDFLLGSLDIAPGMDKVARRLSLIKHPECSTMSGKAI EHLAKQGNRFHFDIK 240

CC sequences of 48-1256 amino acids; a natural amino acid sequence at least

CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino  
CC acids, or 97% identical to a sequence of 242 amino acids, all given in  
CC the specification; or a biologically active or immunogenic fragment of  
CC the isolated polypeptide. The polypeptides and polynucleotides are useful  
CC in diagnosing, treating and preventing diseases or conditions associated  
CC with the decreased expression of protein modification and maintenance  
CC molecules (PMOD), such as cell proliferative diseases (e.g. cancer,  
CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,  
CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,  
CC epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.  
CC endometriosis), developmental, vesicle trafficking disorders, and  
CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also  
CC useful in assessing the effects of exogenous compounds on the expression  
CC of nucleic acid and amino acid sequences of PMOD. The PMOD or its  
CC fragments are useful in screening compounds for effectiveness as agonist  
CC or antagonist of the polypeptides, or in altering the expression of the  
CC target polynucleotide and compounds that specifically bind to or modulate  
CC the activity of the polypeptide. The microarray is useful in monitoring  
CC or measuring protein-protein interactions, drug-target interactions, and  
CC gene expression profiles. This sequence represents a human PMOD protein  
CC of the invention  
XX

SQ Sequence 414 AA;

Query Match 100.0%; Score 414; DB 6; Length 414;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITKTAGVFFPKSKRYEFLRSFNFRHGTLPFLHKIVLGIEISCDPTAAAVNDETGV 60  
|||  
Db 1 MLITKTAGVFFPKSKRYEFLRSFNFRHGTLPFLHKIVLGIEISCDPTAAAVNDETGV 60  
61 LGEAHSOTEVHLLKTGGIVPPAAQOLHRENIRIVOEALASGVSPSDLSAIAITIKPGL 120  
|||  
Db 61 LGEAHSOTEVHLLKTGGIVPPAAQOLHRENIRIVOEALASGVSPSDLSAIAITIKPGL 120  
121 ALSLGVGLSFSIQLVGOLKKEPIPIHMEAHALITRLTNKVEFPFLVLLISGHCILALV 180  
|||  
Db 121 ALSLGVGLSFSIQLVGOLKKEPIPIHMEAHALITRLTNKVEFPFLVLLISGHCILALV 180  
181 QGVSDPFLILGKSLDIAPGMDLKVARRLSLIHQPCSTWSGKAIEHLAKQNRFFEDIK 240  
|||  
Db 181 QGVSDPFLILGKSLDIAPGMDLKVARRLSLIHQPCSTWSGKAIEHLAKQNRFFEDIK 240  
241 PPLHAKKCDSPFSGLOHVTDKIIMKKEEGIEKGQILSSAADIAATVQHTMACHLVNR 300  
|||  
Db 241 PPLHAKKCDSPFSGLOHVTDKIIMKKEEGIEKGQILSSAADIAATVQHTMACHLVNR 300  
301 THRAILFCQKRDLPQNNNAVLVAGSVASNFYIRALBITLTATOCCTLLCPPRLCTDNG 360  
|||  
Db 301 THRAILFCQKRDLPQNNNAVLVAGSVASNFYIRALBITLTATOCCTLLCPPRLCTDNG 360  
361 IMIAMNGIERLRAGILHDIIEGIRYEPKCPGVDISKEVGASIKVPQLKMEI 414  
|||  
Db 361 IMIAMNGIERLRAGILHDIIEGIRYEPKCPGVDISKEVGASIKVPQLKMEI 414

Search completed: November 11, 2005, 03:18:08  
Job time : 120 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2005, 02:14:24 ; Search time 38 Seconds  
(without alignments)  
1048.256 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 414  
Sequence: 1 MLIIITKTAGVFFPKSKRRVY.....DISKEVGESAIKVPOLKWEI 414

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 274

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR\_79: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	Query	
No.	Score	Match length DB ID Description
-----		

No matches found

Search completed: November 11, 2005, 03:22:58  
Job time : 38 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 22:37:04 ; Search time 120 Seconds  
(without alignments)  
1766.673 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 414  
Sequence: 1 MLILTTAGVFPKPSKRRKY.....DISKEVGASIKVPLKMEI 414

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1612378 seqs, 512079187 residues

Word size: 274

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: UniProt\_03:.\*  
1: uniprot\_sprot:.\*  
2: uniprot\_trembl:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	414	100.0	414	2 Q96EV9	Q96ev9 homo sapien

#### ALIGNMENTS

##### RESULT 1

Q96EV9 PRELIMINARY; PRT; 414 AA.  
AC Q96EV9;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
DE OSCEPI1 protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Muscle;  
RX MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haich F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ueidi T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buttefield Y.S.,  
RA Krzywinski M.I., Skalka U., Smalhus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Muscle;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC011904; AA011904.1; -.  
DR MEROPS: M22.004; -.  
DR GO: GO:008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO: GO:008270; F:zinc ion binding; IEA.  
DR GO: GO:006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR000905; Peptidase\_M22.  
DR InterPro: IPR009180; Pept\_M22\_Osialgl.  
DR Pfam: PF00814; Peptidase\_M22; 1.  
DR PIRSF: PIRSF004537; Osialglc\_pptds; 1.  
DR PRINTS: PR00789; OSIALOPTASE.  
DR Prodom: PD002367; Peptidase\_M22; 1.  
DR TRIGRAMS: TRGR00329; gcp; 1.  
SQ SEQUENCE 414 AA; 45122 MW; A536B33F5C6B8BD CRC64;

Query Match 100.0%; Score 414; DB 2; Length 414;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLILTTAGVFPKPSKRRKYEFPLRSFNPHGTLFLHKIVLGITSCDDTAAVVDGTGNV	60
DB	1	MLILTTAGVFPKPSKRRKYEFPLRSFNPHGTLFLHKIVLGITSCDDTAAVVDGTGNV	60
QY	61	LGSAIHSGTEVHLKGTGIIVPPAQQLRENIQRIVDALASASGVSPDSALATTKRGL	120
DB	61	LGSAIHSGTEVHLKGTGIIVPPAQQLRENIQRIVDALASASGVSPDSALATTKRGL	120
QY	121	ALSLGVGLSFLQVLQQLKKPFIPIHMEAHATITRLTNKVEPFLVLLISGCHLALV	180
DB	121	ALSLGVGLSFLQVLQQLKKPFIPIHMEAHATITRLTNKVEPFLVLLISGCHLALV	180
QY	181	QGVSDFLILGKSLIDIAFGDMLDKVARRLSLIKHPECSYSGKAIIEHLAQGRFPHDIK	240
DB	181	QGVSDFLILGKSLIDIAFGDMLDKVARRLSLIKHPECSYSGKAIIEHLAQGRFPHDIK	240
QY	241	PLIHAANKDFSTGQHTYTDKIIMKKEKEBGEIEKQIILSSADIAATVQHTMACHLYKR	300
DB	241	PLIHAANKDFSTGQHTYTDKIIMKKEKEBGEIEKQIILSSADIAATVQHTMACHLYKR	300
QY	301	THRAIFCKORDDLPONNAVIVASGVASNYFIRALAEITLNTAQTCTLCPPLCTDNG	360
DB	301	THRAIFCKORDDLPONNAVIVASGVASNYFIRALAEITLNTAQTCTLCPPLCTDNG	360
QY	361	IMIAWNGIERLRAGLILHIDIEGIRYEPKCPGLVDISKEVGASIKVPLKMEI	414
DB	361	IMIAWNGIERLRAGLILHIDIEGIRYEPKCPGLVDISKEVGASIKVPLKMEI	414

Search completed: November 11, 2005, 03:22:09  
Job time: 120 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2005, 03:16:49 ; Search time 43 Seconds  
(without alignments)  
718.714 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 414  
Sequence: 1 MLILTKTAGVFFKSKRVKRY.....DISKEVEASIKVPLKMEI 414

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 513545 seqs, 74649064 residues

Word size : 274

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- Issued Patents\_AA:\*
- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*
  - 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*
  - 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*
  - 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*
  - 5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*
  - 6: /cgn2\_6/ptodata/1/1aa/Backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	414	100.0	414	4	US-10-067-443-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-10-067-443-2  
; Sequence 2, Application US/10067443  
; Patent No. 6642041  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; Seq ID NO 2  
; LENGTH: 414  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-10-067-443-2

Query Match 100.0%; Score 414; DB 4; Length 414;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLILTKTAGVFFKSKRVKRYEFLRSFNHFGTLFLHKVLGIEISCDPTAAAVDETSNV	60
DB	1	MLILTKTAGVFFKSKRVKRYEFLRSFNHFGTLFLHKVLGIEISCDPTAAAVDETSNV	60
QY	61	LGEAHSOTVHLKTKGIVPPAAOQLRENTQRIVOEALSLASGVSPSLAIAATTIKPGL	120
DB	61	LGEAHSOTVHLKTKGIVPPAAOQLRENTQRIVOEALSLASGVSPSLAIAATTIKPGL	120
QY	121	ALSLGVLSPSLQVLVQLKKPFIPIHMEAHALITRLTNKVEPFLVLLISGHCIALAV	180
DB	121	ALSLGVLSPSLQVLVQLKKPFIPIHMEAHALITRLTNKVEPFLVLLISGHCIALAV	180
QY	181	QGVSDPFLLLGKSLDIAPQMDLDKVARRLSLIKHECSTMSGKAI EHLAKQGNRPHFDIK	240
DB	181	QGVSDPFLLLGKSLDIAPQMDLDKVARRLSLIKHECSTMSGKAI EHLAKQGNRPHFDIK	240
QY	241	PPLHAKNCPSPFTGLQHTVDKTIIMKEKEGIEKQILSSAADIAATVQHTMACHLVKR	300
DB	241	PPLHAKNCPSPFTGLQHTVDKTIIMKEKEGIEKQILSSAADIAATVQHTMACHLVKR	300
QY	301	THRALFCKORDLLPQNNAVLVASGVASNFYIRRALEITLTNATQCTLLCPPPRLCTDNG	360
DB	301	THRALFCKORDLLPQNNAVLVASGVASNFYIRRALEITLTNATQCTLLCPPPRLCTDNG	360
QY	361	IMIAMNGIERLRAGLGIHDIIEGIRYEPKCPGLVDISKVEASIKVPLKMEI 414	
DB	361	IMIAMNGIERLRAGLGIHDIIEGIRYEPKCPGLVDISKVEASIKVPLKMEI 414	

Search completed: November 11, 2005, 03:23:52  
Job time : 43 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2005, 03:17:09 ; Search time 113 Seconds

(without alignments)  
1532.934 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 414  
Sequence: 1 MLITKTAGVFFPKSRKRVY.....DISKEVGASIRVQLKMEI 414

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1867879 seqs, 418409474 residues

Word size : 274

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*

3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*

4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*

6: /cgn2\_6/ptodata/1/pubpaa/PCTOS\_PUBCOMB.pep:\*

7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*

8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*

9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*

10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*

11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*

12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*

13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*

14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*

15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*

16: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*

17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep:\*

18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*

19: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*

21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*

22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	414	100.0	414	US-10-067-443-2	Sequence 2, Appli
2	414	100.0	414	US-10-649-273-2	Sequence 2, Appli
3	414	100.0	414	US-10-651-722-2	Sequence 2, Appli
4	414	100.0	414	US-10-480-988-8	Sequence 8, Appli

#### ALIGNMENTS

RESULT 1  
US-10-067-443-2

Sequence 2, Application US/10067443  
Publication No. US20030082782A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED  
TITLE OF INVENTION: SPINAL CORD, MP-1  
FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
CURRENT FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 2  
LENGTH: 414  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-067-443-2

Query Match 100.0%; Score 414; DB 14; Length 414;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITKTAGVFFPKSRKRVYFNFHFGTLFLHKIVGIETSCDDTAAYVDEKGV 60  
DB 1 MLITKTAGVFFPKSRKRVYFNFHFGTLFLHKIVGIETSCDDTAAYVDEKGV 60

QY 61 LGEAHSQTEVHLKTGQIVPAAQOLHRENIQRIQVQALSASGVSPDLSAATTIKGL 120  
DB 61 LGEAHSQTEVHLKTGQIVPAAQOLHRENIQRIQVQALSASGVSPDLSAATTIKGL 120

QY 121 ALSIGLSTSLQVQOLKRPFIPIHMEHAATITRTNKVEPPVLISGHCILAV 180  
DB 121 ALSIGLSTSLQVQOLKRPFIPIHMEHAATITRTNKVEPPVLISGHCILAV 180

QY 181 QGVSDFLLGKSLDIPGMDLVAVRRLSLIKHPECSTSGKAIEHLAKQGRFHPDIK 240  
DB 181 QGVSDFLLGKSLDIPGMDLVAVRRLSLIKHPECSTSGKAIEHLAKQGRFHPDIK 240

QY 241 PPLHAKNDFSFTGLQHTVTDKIMKKEKEGIEKQILSSADIAATVQHTACHLYKR 300  
DB 241 PPLHAKNDFSFTGLQHTVTDKIMKKEKEGIEKQILSSADIAATVQHTACHLYKR 300

QY 301 THRALLFCQKRDLLPNNNAVLVAGSVAFYIRRALEIITNATQCTLLCPPRLCTDNG 360  
DB 301 THRALLFCQKRDLLPNNNAVLVAGSVAFYIRRALEIITNATQCTLLCPPRLCTDNG 360

QY 361 IMIANGIRLRAGLIDHIDIGIRYEPKPCPLGVDSKEVGASIRVQLKMEI 414  
DB 361 IMIANGIRLRAGLIDHIDIGIRYEPKPCPLGVDSKEVGASIRVQLKMEI 414

RESULT 2  
US-10-649-273-2

Sequence 2, Application US/10649273  
Publication No. US20040043407A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
FILE REFERENCE: D0073 CNT  
CURRENT APPLICATION NUMBER: US/10/649,273  
CURRENT FILING DATE: 2003-08-27  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 10/067,443  
PRIOR FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 2

```

; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-649-273-2

Query Match      100.0%; Score 414; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFPFKSRKRYEFLRSFNFHPTLFLHKIVLGITSCDDTAADVDEGTNV 60
DB 1 MLILTKTAGVFPFKSRKRYEFLRSFNFHPTLFLHKIVLGITSCDDTAADVDEGTNV 60
QY 61 LGEAHSQTEVHLKKGIVPPAAQQLHRENIQRIYQEAALSASGVSPSDSAIATTKRGL 120
DB 61 LGEAHSQTEVHLKKGIVPPAAQQLHRENIQRIYQEAALSASGVSPSDSAIATTKRGL 120
QY 121 ALSAGVGSFSLQVLGOLKKPFIPIHMEBAHALTTRLTKVFPFLVLLISGHCILALV 180
DB 121 ALSAGVGSFSLQVLGOLKKPFIPIHMEBAHALTTRLTKVFPFLVLLISGHCILALV 180
QY 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTWSGKALIEHLAKQGNRFHFDIK 240
DB 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTWSGKALIEHLAKQGNRFHFDIK 240
QY 241 PPLHAKKNCDFSTGLQHTYTDKIIMKEKEBEGIEKQIILSSADIAATVQHTMACHLVVR 300
DB 241 PPLHAKKNCDFSTGLQHTYTDKIIMKEKEBEGIEKQIILSSADIAATVQHTMACHLVVR 300
QY 301 THRALLFCQRDLTPONNAVIVASGVASNFIYRRLAEITLTAOTCTLLCPPRLCTDNG 360
DB 301 THRALLFCQRDLTPONNAVIVASGVASNFIYRRLAEITLTAOTCTLLCPPRLCTDNG 360
QY 361 IMIANGIERLRAGLGIHLHDIEGIRYEPKPLGVDISKVEGASIRVPOLKMEI 414
DB 361 IMIANGIERLRAGLGIHLHDIEGIRYEPKPLGVDISKVEGASIRVPOLKMEI 414

RESULT 3
US-10-651-722-2
; Sequence 2, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bistrol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-651-722-2

Query Match      100.0%; Score 414; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFPFKSRKRYEFLRSFNFHPTLFLHKIVLGITSCDDTAADVDEGTNV 60
DB 1 MLILTKTAGVFPFKSRKRYEFLRSFNFHPTLFLHKIVLGITSCDDTAADVDEGTNV 60
QY 61 LGEAHSQTEVHLKKGIVPPAAQQLHRENIQRIYQEAALSASGVSPSDSAIATTKRGL 120
DB 61 LGEAHSQTEVHLKKGIVPPAAQQLHRENIQRIYQEAALSASGVSPSDSAIATTKRGL 120
QY 121 ALSAGVGSFSLQVLGOLKKPFIPIHMEBAHALTTRLTKVFPFLVLLISGHCILALV 180
DB 121 ALSAGVGSFSLQVLGOLKKPFIPIHMEBAHALTTRLTKVFPFLVLLISGHCILALV 180
QY 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTWSGKALIEHLAKQGNRFHFDIK 240
DB 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTWSGKALIEHLAKQGNRFHFDIK 240
QY 241 PPLHAKKNCDFSTGLQHTYTDKIIMKEKEBEGIEKQIILSSADIAATVQHTMACHLVVR 300
DB 241 PPLHAKKNCDFSTGLQHTYTDKIIMKEKEBEGIEKQIILSSADIAATVQHTMACHLVVR 300
QY 301 THRALLFCQRDLTPONNAVIVASGVASNFIYRRLAEITLTAOTCTLLCPPRLCTDNG 360
DB 301 THRALLFCQRDLTPONNAVIVASGVASNFIYRRLAEITLTAOTCTLLCPPRLCTDNG 360
QY 361 IMIANGIERLRAGLGIHLHDIEGIRYEPKPLGVDISKVEGASIRVPOLKMEI 414
DB 361 IMIANGIERLRAGLGIHLHDIEGIRYEPKPLGVDISKVEGASIRVPOLKMEI 414
```

```

QY 121 ALSAGVGSFSLQVLGOLKKPFIPIHMEBAHALTTRLTKVFPFLVLLISGHCILALV 180
DB 121 ALSAGVGSFSLQVLGOLKKPFIPIHMEBAHALTTRLTKVFPFLVLLISGHCILALV 180
QY 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTWSGKALIEHLAKQGNRFHFDIK 240
DB 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTWSGKALIEHLAKQGNRFHFDIK 240
QY 241 PPLHAKKNCDFSTGLQHTYTDKIIMKEKEBEGIEKQIILSSADIAATVQHTMACHLVVR 300
DB 241 PPLHAKKNCDFSTGLQHTYTDKIIMKEKEBEGIEKQIILSSADIAATVQHTMACHLVVR 300
QY 301 THRALLFCQRDLTPONNAVIVASGVASNFIYRRLAEITLTAOTCTLLCPPRLCTDNG 360
DB 301 THRALLFCQRDLTPONNAVIVASGVASNFIYRRLAEITLTAOTCTLLCPPRLCTDNG 360
QY 361 IMIANGIERLRAGLGIHLHDIEGIRYEPKPLGVDISKVEGASIRVPOLKMEI 414
DB 361 IMIANGIERLRAGLGIHLHDIEGIRYEPKPLGVDISKVEGASIRVPOLKMEI 414

RESULT 4
US-10-480-988-8
; Sequence 8, Application US/10480988
; Publication No. US20050069877A1
; GENERAL INFORMATION:
; APPLICANT: GANDHI, Ameena R.; KABIR, Amy E.;
; APPLICANT: SWARNAKAR, Anita; HARPLA, April J.A.;
; APPLICANT: TRAN, Bao; DUGAN, Brendan M.;
; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;
; APPLICANT: HONCHELL, Cynthia D.; NGUYEN, Danielle B.;
; APPLICANT: LU, Dyung Aina M.; LEE, Ernestine A.;
; APPLICANT: YUE, Henry; FORSYTHE, Ian J.;
; APPLICANT: BARROSO, Ines; RAMKUMAR, Jayalaxmi;
; APPLICANT: GRIFFIN, Jennifer A.; LI, Jona X.;
; APPLICANT: YANG, Junming; THANGAVELU, Kavitha;
; APPLICANT: GIETZEN, Kimberly J.; DING, Li;
; APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.;
; APPLICANT: YAO, Monique G.; CHAWLA, Natinder K.;
; APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
; APPLICANT: LEE, Sally; BECHA, Shanya D.;
; APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;
; APPLICANT: ELLIOTT, Vicki S.; LUO, Wen;
; APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;
; APPLICANT: LU, Yan; ZEBARUDIAN, Yeganeh
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PF-1040 USN
; CURRENT APPLICATION NUMBER: US/10/480,988
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: PCT/US02/19360
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/300,508
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/303,445
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/305,405
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/311,442
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 60/314,821
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/315,992
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 60/378,205
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID No: 7632424CD1  
US-10-480-988-8

Query Match 100.0%; Score 414; DB 17; Length 414;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MLILTKTAGVFPKSKRYEFLRSFNHFGTLFLHKIVLGIETSCDDTAADVDETCNV	60
Db	1	MLILTKTAGVFPKSKRYEFLRSFNHFGTLFLHKIVLGIETSCDDTAADVDETCNV	60
Qy	61	LGEAHSQTEVHLKTGGIVPPAQQLRHNRIQIVQEAISASGVSPSDLSAIAATTIKPGL	120
Db	61	LGEAHSQTEVHLKTGGIVPPAQQLRHNRIQIVQEAISASGVSPSDLSAIAATTIKPGL	120
Qy	121	ALSLGVGLSFSLOLVGOLKKPFIPIHMEAHALTRLTKVFPPLVLLISGHCCLALV	180
Db	121	ALSLGVGLSFSLOLVGOLKKPFIPIHMEAHALTRLTKVFPPLVLLISGHCCLALV	180
Qy	181	QGVSDFLILGKSLDIPGDMLDKVARRLSLIKHPECSTWGGKAIIEHLAKOGNRFHFDIK	240
Db	181	QGVSDFLILGKSLDIPGDMLDKVARRLSLIKHPECSTWGGKAIIEHLAKOGNRFHFDIK	240
Qy	241	PPLHAKKCDPFSFTGLQHVTDKIIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYKR	300
Db	241	PPLHAKKCDPFSFTGLQHVTDKIIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYKR	300
Qy	301	THRAILPCKQDILPQNNNAVVASGVASNFYIRRALIELTNAOTCTLLCPPRLCTDNG	360
Db	301	THRAILPCKQDILPQNNNAVVASGVASNFYIRRALIELTNAOTCTLLCPPRLCTDNG	360
Qy	361	IMIAMNGIERLRAGLGIHDIIEGIRYEPKPLGVDISKVEGASIKVPOLKMEI	414
Db	361	IMIAMNGIERLRAGLGIHDIIEGIRYEPKPLGVDISKVEGASIKVPOLKMEI	414

Search completed: November 11, 2005, 03:25:57  
Job time : 113 secs





QY 221 GYGLYLSAIAIEGLUHIIELEUALIYSGINGIYASNAARGPHEHISPHESAPILIELYS 240  
DB 891 GGTGGGAAAGCATGGAACATTTGGCCAAACAGAAATAGATTTCATTTGACATCAAA 950  
QY 241 PROPROLEUHIISHISALYASNCYSAPPHESERPHETHRGLYLEUGLNIISVALIHR 260  
DB 951 CTTCCCTTGATCATGCTAATAAATTGTGATTTTCTTTACTGGAATTCACACGTTACT 1010  
QY 261 ASPLYSIIIEIEMETLYESGLULYSGIUGIUGIYIIIEGLULYSGIULIIELEUSER 280  
DB 1011 GATAAAATATATATGAAAAAGAAAAAGAGAAAGATTGAGAAAGGGGAAATCTGTCT 1070  
QY 281 SERIAAIAAPPIIEALALATHRVALGINHISTHMEALCYSHISLEUVALLYSARG 300  
DB 1071 TCAGAGAGACATGTGTCGACAGTACGACACATGACATGTCATCTTGTAAGA 1130  
QY 301 THRHSATGALAIIELEUPHESYSGINARGASPLEULEUPROGINAHENALAVAI 320  
DB 1131 ACACATCGGCGTATTCTGTTTGTAGAGAGAGACTTGTACTCAAAATATATGACAT 1190  
QY 321 LEUVALIASERGLYIVAIALASERAPHEIYRIEARGARGALALEUGLIULEU 340  
DB 1191 CTGGTTGATCTGGTGTGTCGCAAGTAACTTCTATATCCGACAGACTGTGAAATTTTA 1250  
QY 341 THRANALATHRGINCYETHREULEUCYSPROPROARGLEUCYETHRASPANGIY 360  
DB 1251 ACAAACGAGACACATGACCTTGTGTGTCTCTCCACGACTATGACCTGTAAATGCG 1310  
QY 361 IIEMETIEALATRPASNGIYIIIEGLUARGLEUARGIAGIYLEUGIYIIIELEUHIISAP 380  
DB 1311 ATTATGATTCATGTAATGTATTTGAAGACATACGCTGCTGGGCAITTTTACATGAC 1370  
QY 381 IIEGLUGIYIIIEARGTYGTIGUBROLYSCYSPROLEUGIYVAILASPIIESEIYSGIULAI 400  
DB 1371 ATAGAAGCATCGCTATGAAACCAAAATGCTCTTGAGTAGACATATCAAAAAGATT 1430  
QY 401 GYGLIUALASERIIELYSVALIPIROGINLEULEUYSMEGLIUIE 414  
DB 1431 GGAGAGGCTTCATTAAGTACCAATTAATAATGAGATA 1472

## RESULT 2

US-09-774-528-177

Sequence 177, Application US/09774528

Patent No. 6743619

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Zhou, Ping

APPLICANT: Goodrich, Ryle

APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod

APPLICANT: Ren, Feiyan

APPLICANT: Zhang, Jie

APPLICANT: Zhao, Qing A.

APPLICANT: Yang, Yonghong

APPLICANT: Xue, Aidong J.

APPLICANT: Wehrman, Tom

APPLICANT: Wang, Jian-Rui

APPLICANT: Wang, Dunrui

APPLICANT: Drmanac, Radoje T.

TITLE OF INVENTION: No. 6743619el Nucleic Acids and

FILE REFERENCE: 802

CURRENT APPLICATION NUMBER: US/09/774,528

NUMBER OF SEQ ID NOS: 441

SOFTWARE: pc\_fl\_genes Version 2.0

SEQ ID NO 177

LENGTH: 1416

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (205)..(1305)  
US-09-774-528-177  
Alignment Scores:  
Pred. No.: 0  
Score: 340.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 82.13%  
DB: 4  
Gaps: 0  
US-10-649-273-2 (1-414) x US-09-774-528-177 (1-1416)  
QY 75 THRGLYIYIEVALIPIROPROALAIAGINGLNIIEHISARGIUBENIIEGINARGIIE 94  
DB 283 ACAGGTGGATTTCTTCTCCAGACGCTCAACACCTTCACAGAAATATTCACAGATA 342  
QY 95 VALINGIUALAUSERALASERGLYVALISERPROSERAPLESERAIIEALATHR 114  
DB 343 GTACAAAGAGCTTTCTGTCAGTAGAGTCTTCCAAAGTACCTTCACGCAATTCGCACT 402  
QY 115 THRILEYSPROGLYLEUALALEUSERLEUGIYVAILYIEUSERPHESERIEUGLNU 134  
DB 403 ACCATTAACACAGACTTGTCTTAAGCTGGAGTGGCTTATCATTTACTTACACACTG 462  
QY 135 VALIGIULNUYLSYSPROPEIIEPROIIEHISMETGLUALAHSALALEUTHR 154  
DB 463 GTTGACAGATTAAAGCATTCATTCATTCATATATGAGGCTCATGCACTTACT 522  
QY 155 IIEARGLEUTHRANLYSVALIUIPHEPROPHLEUVALLEUUIIESEGLYGIYHIS 174  
DB 523 ATTAGGTGACATATAAGTAGAATTTCTTTTATGTTCTTTGATTTCTGAGGTAC 582  
QY 175 CYELEULEUALALEUVALINGIYVALISERASPHLEULEULEUGIYLSERLEUASP 194  
DB 583 TGTCTGTGTCATTAGTCAAGAGATTTCAGATTTTCGTCTTGTGAAAGTCTTTGAC 642  
QY 195 IIEALAPIROGLYASPHLEUASPLYSVALIARGARGLEUSERLEUIELYSHIISPRO 214  
DB 643 ATGACACCAAGTACATGCTTGACAGGTGCAAGAAAGCACTTTTAAATTAACATCA 702  
QY 215 GLUCYSETHRMESESERGLYGLYLSAIAIEGLUHIIELEUALIYSGINGIYASNARG 234  
DB 703 GAGTGTCCACACATGAGTGTGGAAAGCCATGAAACATTTGGCAAAAGAAATAGA 762  
QY 235 PHEHISPHESAPILIELYSPPROLEUHIISHISALYASNCYSAPPHESERPHETHR 254  
DB 763 TTTCAATTTGACATCAAACTCCCTTCATCATGCTAATAAATGTGATTTTCTTTACT 822  
QY 255 GYILEUGLNIISVALIATHRASPILIEIEMETLYESGLULYSGIUGIUGIYIIIEGLU 274  
DB 823 GGACTTCACACGTTACTGATTAATAATATGAAAAAGAAAAAGAAAGGATTTGAG 882  
QY 275 LYEGIYGINIIELEUSERSERALASAPPIIEALALATHRVALGINHISTHMEALIA 294  
DB 883 AAGGGGAAATCTCTGTTTACGACGACATTTCTCCACAGACACATGCA 942  
QY 295 CYSHISLEUVALLYSARGTHRHISARGALAIIELEUPHESYSGINARGASPLEULEU 314  
DB 943 TGTATCTTGGAAGAAACACACATCGGCTATCTGTTTGTAGACAGAGACTTGTTA 1002  
QY 315 PROGINAHENALAVAILLEUVALIASERGLYIVAIALASERAPHEIYRIEARG 334  
DB 1003 CTTCAAAATATATGACAGTACGTTGTCATCTGTGTGCGAAGTAACCTTATATCCGC 1062  
QY 335 ARGALALEUGLIULEUTHRANALATHRGINCYETHREULEUCYSPROPROARG 354  
DB 1063 AGAGCTTCGAAATTTTAAACAAACGACACAGTCACTTTGTGTCTCTCCGCA 1122  
QY 355 LEUCYETHRASPANGIYIIEMETIEALATRPASNGIYIIIEGLUARGLEUARGIAGIY 374  
DB 1123 CTATGACATGATATGCAATTAATGATTCATGCAATGATGATTAAGAACTACGTGCTGC 1182

QY 375 LeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProIysCysProLeuGlyVal 394  
 |||||  
 Db 1183 TTGGGCATTTTACATGACATGAGGCAATCCGCTATGAACCAAAATGCTCTTGGAGTA 1242  
 |||||  
 QY 395 AspIleSerIysGluValGlyGluIleSerIleIysValProGlnLeuIysMetGluIle 414  
 |||||  
 Db 1243 GACATATCAAAAGAAAGTTGGAGAGCTTCCATTAAGTACACACATTTAAATAATGAGATA 1302  
 |||||

Search completed: November 11, 2005, 07:21:00  
 Job time : 298 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 11, 2005, 03:18:15 ; Search time 887 Seconds  
(without alignments)  
2762.991 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 414  
Sequence: 1 MLILTKTAGVFFKPSKRYV.....DISKEVGASIVPOLKMEI 414

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Word size: 274

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US10649273/runat\_02112005\_031640\_16863/app\_query.fasta\_1.583  
-DB=N/Geneseq\_16Dec04 -QFMT=fastap -SUFFIX=oligp2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCCALIGN=200 -THR SCORE=quality -THR MIN=274 -ALIGN=15 -MODS=LOCAL  
-OUTFMT=plco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US10649273 @CGN 1.1 708 @runat\_02112005\_031640\_16863 -NCPU=6 -ICPU=3  
-NO\_MMAR -LARGEOQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :  
1: Geneseq\_16Dec04:\*  
2: geneseqn1980s:\*  
3: geneseqn1990s:\*  
4: geneseqn2000s:\*  
5: geneseqn2001as:\*  
6: geneseqn2001bs:\*  
7: geneseqn2002as:\*  
8: geneseqn2002bs:\*  
9: geneseqn2003as:\*  
10: geneseqn2003bs:\*  
11: geneseqn2003cs:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	414	100.0	2058	ABA93268	ABA93268 Human O-s
2	414	100.0	2197	AB876635	AB876635 DNA encod
3	414	100.0	2572	ABT23207	ABT23207 Human pro
4	340	82.1	1416	ABX70950	ABX70950 Novel hum

ALIGNMENTS

RESULT 1	
ID	ABA93268 standard; cDNA; 2058 BP.
AC	ABA93268;
DT	19-APR-2002 (first entry)
DE	Human O-sialoglycoproteinase-like protein encoding cDNA SEQ ID NO.1.
XX	Human; O-sialoglycoproteinase-like protein; OSGPUP; enzyme; gene; ss.
XX	OS
XX	Homo sapiens.
OS	
FT	Key
FT	location/Qualifiers
FT	CDS
FT	110..1354
FT	/*tag= a
XX	/product= "O-sialoglycoproteinase-like protein"
PN	CN1318550-A.
XX	
PD	24-OCT-2001.
XX	
PP	19-APR-2000; 2000CN-00106834.
XX	
PR	19-APR-2000; 2000CN-00106834.
XX	
XX	(SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.
XX	
PI	Mao Y, Xie Y;
XX	
DR	WPI; 2002-115090/16.
DR	P-PSDB; ABB05481.
XX	
PT	O-sialoglycoproteinase-like protein and encoding polynucleotide, useful
XX	for diagnosing, preventing and treating related diseases.
XX	
PS	Claim 5; Page 29-30 (disclosure); 38pp; Chinese.
XX	
XX	The present sequence encodes human O-sialoglycoproteinase-like protein
CC	(OSGPUP). The present invention also describes: (1) the preparation of
CC	the OSGPUP protein; (2) applying the OSGPUP protein in diagnosis; (3) the
CC	prevention and/or treatment of related diseases; (4) utilizing the OSGPUP
CC	protein in screening its agonist, excitomotor and inhibitor and preparing
CC	an antibody against the OSGPUP protein; and (5) the use of the OSGPUP
CC	polynucleotide sequences, proteins, agonists, excitomoters, inhibitors
CC	and antibodies in treating diseases related to the abnormal OSGPUP gene
XX	and in preparing the medicine composite for the treatment
XX	
SQ	Sequence 2058 BP; 637 A; 400 C; 410 G; 611 T; 0 U; 0 Other;
Alignment Scores:	
Pred. No.:	0
Score:	414.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	6
	Gaps: 0
US-10-649-273-2 (1-414) x ABA93268 (1-2058)	
QY	1 MetLeuIleLeuThrIysThraIaGlyValPhePheIysProSerIysArgIysValTyr 20
DB	110 ATCTAATCTTGACTGAAGACTGACGAGATTTTAAACCACTAAAGAAAGTTAT 169
QY	21 GlnPheIenArgSerPheIenPheIserPrGlyThreIenPheIenIcIysIleValIcu 40
DB	170 GAATTTTAAAGATTTAATTTTCATCCGAAACATATTTCTTCATAAAATAGTATTG 229
QY	41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60

```

Db      230 GGAATTGAAACAGTGTGATGATACAGAGCTGCTGTGTGATGAAACTGGAATGTG 289
Qy      61 LeuGIyGluAaIleHISerGIInThrGIuValHISleuLyThrGIyLIleValPro 80
Db      290 TTGGGAGAGACAAATCATTTCCAAACCTGAAGTTCAATTTAAAAACAGTGGATGTTCTCT 349
Qy      81 ProHlaaIaGIInGlnLeuHISArgIuAenIIeGIInArgIleValGIInGluAaleuSer 100
Db      350 CCAGCAGCTCAACACCTTCACAGAGAAATATTTCAAGATAGTACAGAGAGCTCTTTCT 409
Qy      101 AlaSerGIyValSerProSerAspLeuSerAlaIleAlaThrThrIleLySProGIyLeu 120
Db      410 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATTGCACTACCAATAAACCAGAGACT 469
Qy      121 AlaLeuSerLeuGIyValGIyLeuSerPheSerLeuGlnLeuValGIyGlnLeuLyS 140
Db      470 GCTTTAAGCTGTGGAGTGGCTTATCATTTAGCTTACAGCTGGTAGAGAGATTAAAAAG 529
Qy      141 ProPheIleProIleHISMetGIuAaHISaIleuThrIleArgLeuThrAsnLyS 160
Db      530 CCATTCATTTCCATTCATCATATGAGAGCTCATGACCTTACTATTAAGTTGACCAATPAA 589
Qy      161 ValGIuPheProPheLeuValLeuLeuIleSerGIyGIyHISCySLeuLeuAlaLeuVal 180
Db      590 GTAGAAATTTCTTTTGTAGTCTTTTGAATTTCTGGAGGTCACTGCTGTTGGCATTTAGT 649
Qy      181 GInGIyValSerAspPheLeuLeuLeuGIyLySLeuAspIleAlaProGIyAspMet 200
Db      650 CAAGAGATTTTCAGATTTTCTGCTTCTTGAAAGTCTTTGGAACATGACACAGGTGACATG 709
Qy      201 LeuAspLySValAlaArgArgLeuSerLeuIleLyHISProGIuCySerThrMetSer 220
Db      710 CTGTGCAAGGTGGCAGAGAGACTTTCTTAAATAAAATCCAGAGAGCTCCACCATGAGAT 769
Qy      221 GIyGIyLySAlaIleGIuHISleuAlaLySGlnGIyASnArgPheHISpHeAspIleLyS 240
Db      770 GGTGGGAGAGCCATTAACATTTGGCCAAACAGAAATAGATTTATTTTGACATCMAA 829
Qy      241 ProProLeuHISHisAlaLySAsnCyAspPheSerPheThrGIyLeuGlnHISValThr 260
Db      830 CCTCCCTTGCAATCATCTAAATAATGTGATTTTCTTTTACTGAGACTTCAACACGTTACT 889
Qy      261 AspLySIIleIleMetLySlySGluLySGluGIyLIleGIuLySGIyGlnIleLeuSer 280
Db      890 GATTAATAATATATGAAAAAGGAAAAAGAGATTAAGAGAGGGGCAAAATCTGTCT 949
Qy      281 SerAlaIleAspIleAlaIleThrValGIInHISThrMetAlaCySHISleuValLySArg 300
Db      950 TCAGCAGCAGACATTTCTGCCACAGTACAGACACACATGAGCTCATTTGTGAAAGA 1009
Qy      301 ThrHISArgAlaIleLeuPheCyLySGlnArgAspLeuLeuProGIInAsnAlaVal 320
Db      1010 ACRATCGGGCTATTTCTGTTTCTTAACAGACAGACTGTGTACTCAAAATAATGACGTA 1069
Qy      321 LeuValAlaSerGIyGIyValAlaAspAsnPheTyrlleArgArgAlaLeuGIuIleLeu 340
Db      1070 CTGGTTCATCTGGTGGTGTCCGACAGTAATTTCTAATACCGCAGAGCTCTGGAATTTTA 1129
Qy      341 ThrAsnAlaIleThrGlnCySerThrLeuLeuCySerProProArgArgLeuCySThrAspAsnGIy 360
Db      1130 ACAAAAGCAGACAGTGCATTTGTGTGTCTCTCCACAGCTATGCACTGATAATGAGC 1189
Qy      361 IleMetIleAlaIleArgAsnGIyIleGIuArgLeuArgAlaGIyLeuGIyIleLeuHISAsp 380
Db      1190 ATTTATGATTCATGATGATGATTTGAAAGCTACAGTCTGCTGGCATTTTATCAGAGAC 1249
Qy      381 IleGIuGIyIleArgTyrgIuProLyCySProLeuGIyValAspIleSerLySLeuVal 400
Db      1250 ATAGAAAGGATCGGCTATGAAACAAATGTCTCTTGAGATAGCATATCAAAAAGAGTT 1309
Qy      401 GIyGIuAlaSerIleLySValProGIInLeuLyMetGIuIle 414

```

```

Db      1310 GGAGAGCTTCATTAAGTACCAATTAATAAATGAGATA 1351
RESULT 2
AB576635
ID      AB576635 standard; DNA; 2197 BP.
XX
AC      AB576635;
XX
DT      11-DEC-2002 (first entry)
XX
DE      DNA encoding novel human metalloprotease MPI.
XX
XX      Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;
KW      motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
KW      reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
KW      genital wart; metabolic disorder; premature puberty; Kallman syndrome;
KW      Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
KW      Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
KW      liver disease; renal disease; immune disorder; rheumatoid arthritis;
KW      acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
KW      emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
KW      neurological disorder; gene; ds.
XX
XX      Homo sapiens.
XX
XX      PN      WO200272751-A2.
XX
XX      PD      19-SEP-2002.
XX
XX      PF      05-FEB-2002; 2002WO-US003353.
XX
XX      PR      05-FEB-2001; 2001US-0266518P.
XX      PR      10-APR-2001; 2001US-0282814P.
XX
XX      PA      (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX      PI      Chen J, Feder J, Nelson TC, Duclos F, Kryetok S;
XX      WPI; 2002-723329/78.
XX      DR      P-PSDB; AB596478.
XX
XX      PT      New isolated nucleic acid encoding MP-1 protein, useful for preventing,
PT      treating, or ameliorating diseases associated with aberrant
PT      metalloprotease activity, e.g. immune, metabolic, inflammatory and
PT      neurological disorders.
XX
XX      PS      Claim 1; Fig 1A-C; 473bp; English.
XX
XX      CC      The invention describes an isolated nucleic acid molecule (1) encoding a
CC      metalloprotease (MP-1). (1) is useful for preventing, treating, or
CC      ameliorating a medical condition, particularly an immune disorder, an
CC      aberrant glutamate transport or motor neuron disorder, such as
CC      amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like
CC      condition. The compositions and methods are also useful for diagnosing,
CC      prognosticating, treating, ameliorating and/or treating disorders
CC      associated with MP-1 activity, e.g. diabetes, cancer, reproductive
CC      disorders (e.g. Klinefelter's syndrome, genital wart, or germinal cell
CC      aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,
CC      or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,
CC      Parkinson's disease, Huntington's disease or Tourette syndrome), liver
CC      and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis
CC      or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic
CC      fibrosis) and vascular, inflammatory and neurological disorders (e.g.
CC      Alzheimer's disease or Parkinson's disease). This sequence represents a
CC      metalloprotease MPI polynucleotide
XX
XX      SO      Sequence 2197 BP; 681 A; 441 C; 439 G; 636 T; 0 U; 0 Other;

```

Alignment Scores: 0  
 Pred. No.: 414.00  
 Score: 100.00%  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Length: 2197  
 Matches: 414  
 Conservative: 0  
 Mismatches: 0

Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
US-10-649-273-2 (1-414) x ABS76635 (1-2197)

QY 1 MetLeuLeuThrThyThraGlyValPhePheLeuProSerIysArgIysValTyr 20  
Db 231 ATGCTAACTTGTACCTAGACAGCTGAGAGTCTTTTAAACCAACCAAGAAAGTTTAT 290

QY 21 GluPheLeuArgSerPheAsnPhelIspProGlyThrLeuPheLeuHisIysIleValLeu 40  
Db 291 GAATTTTAAAGAGTTTAAATTTTCTCTGGAACCTATTTCTTCAATAAATAGATTG 350

QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60  
Db 351 GGAATTTGAACTAGTTGTGATGATACAGACGCTGCTGGTGGATGAATGAACTGGAATGTG 410

QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyGlyIleValPro 80  
Db 411 TTGGAGAGAGCAATTCATTTCCCAACCTGAAGTTTCAATTTAAAAACAGGTGGATTTCTCT 470

QY 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
Db 471 CCAAGAGCTCAACAGCTTCAACAGAAATATTCAACGAATAGTACAGAGAGCTTTTCT 530

QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
Db 531 GCCAGTGGAGTCTCTCCAAGTGACCTCTCAGCAATTTGCAATCACTAATAACAGAGACTT 590

QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysVal 140  
Db 591 GCTTTAAGCTTGGGAGTGGCTTATCATTTACTTACAGCTGTAGAGACAGTTAAAAAG 650

QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
Db 651 CCATTCATTTCCATTCATTCATTCAGAGGCTCATGACCTTACTATTAGTTGACCAATAAA 710

QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
Db 711 GTAGAAATTTCTTTTGTAGTTCTTTGATTTCTGAGAGTCACTGTCTGTGGCATTAAGTT 770

QY 181 GlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
Db 771 CAAGAGATTTCAAGATTTCTGCTCTTGGAAAGTCTTTGAGCATAGCAACAGGTGACATG 830

QY 201 LeuAspLysValAlaArgArgLeuSerIleLysIlePProGluCysSerThrMetSer 220  
Db 831 CTTACCAAGGTGGCAAGAACTTTCTTTAATMAAACCAAGAGTGTCCACCATGAGT 890

QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
Db 891 GGTGGAGAGCAATTCATTCATTTGGCCAAACAGAAATAGATTTCATTTTACATCAAA 950

QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
Db 951 CCTCCCTTGATCATAGTAAAAATGATGATTTTCTTTTACCTGAGCTTCAACAGCTTACT 1010

QY 261 AspLysIleIleMetLysLysGlnLysGluGluGlyIleGlnLysGlyGlnIleLeuSer 280  
Db 1011 GATMAAATTAATMAATGAAGAAAGCAAGAGATTTGAGAGAGGGCAATTCCTGTCT 1070

QY 281 SerAlaIleAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
Db 1071 TCAGCAGAGACATTTGTCTGCACACATACACACAAATGCAATGTCATTTGTGAAAAA 1130

QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320  
Db 1131 ACAATCGGGCTATTCTCTTTTGTAAAGCAGAGACTTCTTACCTCAAAATATATGACAGTA 1190

QY 321 LeuValAlaSerGlyValAlaAlaSerAsnPhelYrIleArgArgAlaLeuGluIleLeu 340  
Db 1191 CTGGTTCATCTGGGTGTGTGCAAGTAATCTTATATCCGACAGCTCTGCAAAATTTTA 1250

QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaArgLeuCysThrAspAsnGly 360  
Db 1251 ACAACCCAAACACAGTGCACCTTTGTGTGTCTCTCTCCAGACTATACATGATTAATGGC 1310

QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuAlaGlyLeuGlyIleLeuHisAsp 380  
Db 1311 ATTATGATTCATGAGATGATGATTAATAAGACTAGCTGCTGGCTGGCATTTTACATGAC 1370

QY 381 IleGluGlyIleArgTyrgluProLysCysProLeuGlyValAspIleSerLysGluVal 400  
Db 1371 ATGAGAGCATCCGCTATGAAACCAAAATGCTCTTGGAGTAGACATATCAAAAGAGTT 1430

QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
Db 1431 GGAGAAAGCTTCATTAATAAGTACACCAATTAATAATGAGAGATA 1472

RESULT 3  
ABT23207  
ID ABT23207 standard; DNA; 2572 BP.  
XX  
AC ABT23207;  
XX  
DT 01-MAY-2003 (first entry)  
XX  
DE Human protein modification + maintenance molecule DNA SEQ ID No 36.  
XX  
CYCOSTATIC; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;  
KW cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;  
KW anticancer; hepatotropic; gynaecological; antibacterial; virucide;  
KW protozoacide; antiparasitic; cell proliferative disease; PMOD;  
KW protein modification and maintenance molecule; immunogenic fragment;  
KW cancer; autoimmune; inflammatory disease; neurological disorder;  
KW gastrointestinal; developmental; vesicle trafficking disorder; infection;  
KW protein-protein interaction; drug-target interaction;  
KW gene expression profile; human; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN MO2003000844-A2.  
XX  
PD 03-JAN-2003.  
XX  
PF 18-JUN-2002; 2002MO-US019360.  
XX  
PR 22-JUN-2001; 2001US-0300508P.  
XX  
PR 06-JUL-2001; 2001US-0303445P.  
XX  
PR 13-JUL-2001; 2001US-0305405P.  
XX  
PR 09-AUG-2001; 2001US-0311442P.  
XX  
PR 24-AUG-2001; 2001US-0314821P.  
XX  
PR 29-AUG-2001; 2001US-0315922P.  
XX  
PR 03-MAY-2002; 2002US-0378205P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
XX Gandhi AR, Kable AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BM;  
PI Warren BA, Ison CH, Honchel CD, Nguyen DB, Lu DAM, Lee BA, Yue H;  
PI Forsthe ID, Barroso I, Rankumar J, Griffith JA, Li JX, Yang J, McG;  
PI Thangavelu K, Gietzen KJ, Ding L, Baughn MK, Borowsky ML, Yao S;  
PI Walla NK, Mason PM, Gunturajan R, Lee S, Becha SD, Lee SY, Tran UK;  
PI Billicot VS, Luo W, Sprague WW, Tang YT, Lu Y, Zebajadian Y;  
XX  
DR WPI: 2003-184039/18.  
DR P-PSDB; AB026654.  
XX  
XX  
PT New isolated human PMOD polypeptide and polynucleotide, useful for  
PT diagnosing, treating and preventing diseases or conditions associated  
PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and  
PT infections.  
XX  
PS Claim 91; Page 211; 225pp; English.  
XX  
CC The invention relates to an isolated polypeptide comprising: any of 28  
sequences of 48-1256 amino acids; a natural amino acid sequence at least

CC 90% identical to the 28 amino acid sequences, 94% identical to a sequence  
 CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino  
 CC acids, or 97% identical to a sequence of 242 amino acids, all given in  
 CC the specification; or a biologically active or immunogenic fragment of  
 CC the isolated polypeptide. The polypeptides and polynucleotides are useful  
 CC in diagnosing, treating and preventing diseases or conditions associated  
 CC with the decreased expression of protein modification and maintenance  
 CC molecules (PMD), such as cell proliferative diseases (e.g. cancer,  
 CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,  
 CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,  
 CC epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.  
 CC endometriosis), developmental, vesicle trafficking disorders, and  
 CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also  
 CC useful in assessing the effects of exogenous compounds on the expression  
 CC of nucleic acid and amino acid sequences of PMD. The PMD or its  
 CC fragments are useful in screening compounds for effectiveness as agonist  
 CC or antagonist of the polypeptides, or in altering the expression of the  
 CC target polynucleotide and compounds that specifically bind to or modulate  
 CC the activity of the polypeptide. The microarray is useful in monitoring  
 CC or measuring protein-protein interactions, drug-target interactions, and  
 CC gene expression profiles. This polynucleotide sequence represents the DNA  
 CC encoding a human PMD protein of the invention

XX Sequence 2572 BP; 780 A; 489 C; 509 G; 794 T; 0 U; 0 Other;

# Alignment Scores:

Pred. No.: 0 Length: 2572  
 Score: 414.00 Matches: 414  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 8 Gaps: 0

US-10-649-273-2 (1-414) x ABT23207 (1-2572)

QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLeuProSerLysArgLysValTyr 20  
 Db 144 ATGCTAATCTTGACATAGACTGACAGAGCTTTTAAACATCAAAAAGAAAGTTTAT 203  
 QY 21 GluPheLeuArgSerPheAsnDheHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
 Db 204 GAATTTTAAAGAAATTTTAAATTTTCACTGGAACACTATTTCTTATATAAATAGATTTG 263  
 QY 41 GlyIleGluThrSerCysAspAspPheAlaAlaValValAspGluThrGlyAsnVal 60  
 Db 264 GGAATTTGAAACTGATGATGATATACAGACGCTGCTGTGATTAACCTGAAATATGTG 323  
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyIleValPro 80  
 Db 324 TTGGAGAAAGCAATACATTCCTCCAACTGAAGTTCATTTTAAAAACAAGTGGATTTCTT 383  
 QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 Db 384 CCAGACAGCTCAACAGCTTCACAGAGAAATATTTCAAGATATAGTACAAAGAACTCTTTCT 443  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
 Db 444 GCCAGTGAAGTCTCTCCCAAGTACCTCTCAGCAATTCGCAATCAATAAACAGAGACTT 503  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140  
 Db 504 GCTTTAAGCTTGAGAGTGGAGCTTATATATTAGCTTCAAGCTGTATAGACAGTTAAAAAG 563  
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
 Db 564 CCATTCAATCCCATTCATCATATGAGGCTCAATGCACTTATAGTTGTTGACCAATATAA 623  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisGlySerLeuLeuAlaLeuVal 180  
 Db 624 GTAGAAATTTCTTTTATCTTTTATGATTTCTGAGGACACATGCTGTGTGGCAATTAAGT 683  
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200

Db 684 CAAAGAGTTTCAGATTTTCTGCTTCTGGAAAAGCTTTTGGACATACAGAGTGACATG 743  
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220  
 Db 744 CTTCAGAAAGTGGCAAGAGACTTCTTTTAAATAAATCCACAGAGTCTCCACATAGT 803  
 QY 221 GlyIleValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
 Db 804 GGTGGGAAAGCCATAGAAACATTTGGCCAAACAGAAATAGATTTCTTTTGAATCAATCAA 863  
 QY 241 ProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGlyLeuGlnHisValThr 260  
 Db 864 CCCCCCTTGATATATGCTAAAAATTTGATTTTCTTTTATCTGAACTTCAACACGTACT 923  
 QY 261 AspLysIleIleMetLysLysGlyLysGluGluGlyIleGluLysGlyGlnIleLeuSer 280  
 Db 924 GATTAATAATATATGAAAAAGAAAAAGAAAGATTTAGAGAGGGGCAAAATCTGTCT 983  
 QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
 Db 984 TCAGACAGACAGATTCGTCCACAGTACAGACACATATGCGATGTCTTGTGAAGAAGA 1043  
 QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320  
 Db 1044 ACACATCGGGCTATTTCTGTTTGTATAGCAGACAGACTTGTACCTCAAAATATATGACGTA 1103  
 QY 321 LeuValAlaSerGlyValAlaAspAspPheTyrIleArgArgAlaLeuGluIleLeu 340  
 Db 1104 CTGTGTGACATCTGT 1163  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360  
 Db 1164 ACAAGGCAACACAGAGCACTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1223  
 QY 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 Db 1224 ATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1283  
 QY 381 IleGluGlyIleArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 400  
 Db 1284 ATGAAAGGCAATCCGCTATGAAACAAATATGCTCTTGTGAATGACATATCAAAAGAGTT 1343  
 QY 401 GlyIleAlaSerIleLysValProGlnLeuLysMetGluIle 414  
 Db 1344 GGAAGAGCTTCCATTAAGATACCAATTTAAAAATGAGATTA 1385

RESULT 4  
 ID ABX70950 standard; cDNA; 1416 BP.  
 AC ABX70950;  
 XX 05-MAR-2003 (first entry)  
 DE Novel human cDNA sequence #175.  
 XX Human; gene; ser. nervous system disorder; peripheral neuropathy;  
 KW Huntington's disease; amyotrophic lateral sclerosis; haemophilia;  
 KW neurodegenerative disease; Parkinson's disease; Alzheimer's disease;  
 KW autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;  
 KW insulin-dependent diabetes mellitus; anaemia; chromocyclopaenia; wound;  
 KW ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;  
 KW fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;  
 KW coagulation disorder; cancer; tumour; inflammatory disease; septic shock;  
 KW Crohn's disease; anaphylaxis; proliferation; chemotactic;  
 KW differentiation; stem cell growth factor; haematopoiesis; chemokinetic;  
 KW haemostatic; antiinflammatory; expressed sequence tag; EST.  
 XX Homo sapiens.  
 OS  
 PN MO200281731-A2.  
 XX  
 PD 17-OCT-2002.

XX 29-JAN-2002; 2002WO-US001222.  
 XX  
 XX 30-JAN-2001; 2001US-00774528. (HYSE-) HXSEO INC.  
 XX (GOOD/) GOODRICH R W.  
 XX  
 XX Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao Q, Ren F;  
 PI Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RR;  
 XX WPI; 2003-058563/05.  
 XX  
 PT Novel polypeptide useful for treating neurodegenerative diseases, myeloid  
 PT or lymphoid cell disorders, bone disorders, mechanical and traumatic  
 PT disorders, coagulation disorders, and inflammatory diseases.  
 XX  
 PS Claim 1; Page: 612pp; English.

CC This invention relates to the cDNA sequences encoding an isolated novel  
 CC human polypeptide. The protein encoded by the nucleic acid of the  
 CC invention is useful for treating central and peripheral nervous system  
 CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic  
 CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,  
 CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus  
 CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)  
 CC or myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopenia)  
 CC or wounds, ulcers, burns, bone disorders (e.g. osteoporosis,  
 CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head  
 CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;  
 CC bacterial, viral or fungal infections; allergic conditions such as  
 CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);  
 CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's  
 CC disease, anaphylaxis). The protein may be used to inhibit the growth,  
 CC infection or function of infectious agents such as bacteria, fungi,  
 CC viruses, or to effect bodily characteristics, biorhythms or circadian  
 CC cycles of rhythms. The protein may also have  
 CC proliferation/differentiation, stem cell growth factor, haematopoiesis  
 CC regulation, immune stimulating or suppressing, chemotactic/chemokinetic,  
 CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory  
 CC activities. The cDNA sequences of the invention are useful for expressing  
 CC recombinant protein for analysis. The present sequence represents a novel  
 CC human cDNA sequence of the invention, this sequence is an expressed  
 CC sequence tag (EST) and was identified using subtractive hybridisation  
 XX  
 XX Sequence 1416 BP; 441 A; 280 C; 273 G; 422 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 0 Length: 1416  
 Score: 340.00 Matches: 340  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 82.13% Indels: 0  
 DB: Gaps: 0

US-10-649-273-2 (1-414) x ABX70950 (1-1416)

QY 75 ThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIle 94  
 DB 283 ACAGGTGGATGTCTCCAGCAAGCTCAACAGCTTCAAGAGAAATATTCACGAAATA 342  
 QY 95 ValGlnGlnIleAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThr 114  
 DB 343 GTACAAAGAGCTCTTCTGCGAGGTGAGTCTCTCCAAAGTACCTCTCAGCAATTCACACT 402  
 QY 115 ThrIleYsProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeu 134  
 DB 403 ACCATTAAACAGAGACTTCTTAAGCTTGAAGTGGGCTTATCATTTAGCTTACAGCTG 462  
 QY 135 ValGlyGlnLeuYsYsProPheIleProIleHisIleMetGluAlaHisIleAlaLeuThr 154  
 DB 463 GTAGGACAGTTAAAGAGCCATTCATTCATTCATCATATGAGAGCTCATGCACTTACT 522

QY 155 IleArgLeuThrAspLeuValGluPheProPheLeuValLeuLeuIleSerGlyGlyHis 174  
 DB 523 ATTAGGTGACCAATAAAGTAGAATTCCTTTTCTAGTCTTTGATTTCTGAGAGCTCAC 582  
 QY 175 CysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlySerLeuAsp 194  
 DB 583 TGTCTGTGGCATTAGTTCAGAGAGTTCAGATTTTCTGCTTCTGGAAAGCTTTTGAC 642  
 QY 195 IleAlaProGlyAspMetLeuAspLeuValAlaArgArgLeuSerLeuIleLeuHisPro 214  
 DB 643 ATAGCACACAGTGCATGCTTGCACAGGTGGCAAGAGACTTTTATTAATTAACATCCA 702  
 QY 215 GluYsSerThrMetSerGlyGlyValAlaIleMetLeuSerGlnGlnGlyValAsnArg 234  
 DB 703 GAGTCTCCACCATGAGTGTGGAAAGCATAGAACATTTGGCCAAACAGAAATAGA 762  
 QY 235 PheHisPheAspIleYsProProLeuHisIleAlaIleValAsnGlyAspPheSerPheThr 254  
 DB 763 TTTCATTTTGACATCAAACTCCCTTGCAATGCTAATAAATGTGTATTTCTTTTACT 822  
 QY 255 GlyLeuGlnHisValThrAspLeuIleIleMetLeuSerGlnGlnGlyValGlnGly 274  
 DB 823 GAGCTTCAACACCTTACTGATTAATATATATGAAAAAGAAAGAAAGGATTTAG 882  
 QY 275 LysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAla 294  
 DB 883 AAGGGCAATCTGTTCTTACAGACAGACATTTGCTCCACAGTACGACACATGGCA 942  
 QY 295 CysHisLeuValIysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeu 314  
 DB 943 TGTCACTTGTGAAAAAGAACACATCGGCTATTTCTGTTGTAAACAGACAGACTGT 1002  
 QY 315 ProGlnAspAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArg 334  
 DB 1003 CTTCAAAATATATACAGTACTGTTGATCTGGTGTGCCAAGTAACTTCTATATCCG 1062  
 QY 335 ArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArg 354  
 DB 1063 AGAGCTCTGGAAATTTTAAACAAACGACACAGTGCCTTTGTTGTCTCTCCACGA 1122  
 QY 355 LeuCysThrAspAsnGlyIleMetIleAlaIleThrAsnGlyIleGluArgLeuAlaGly 374  
 DB 1123 CTATGACATGATATATGCAATTATGATTCATGAAATGTAATTGAAAGACTTACGGCTGGC 1182  
 QY 375 LeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlnProYsCysProLeuGlyVal 394  
 DB 1183 TTGGGCAATTTTACATGACATAGAAAGCATCCGCTATGAAACCAAAATGTCTCTTGAGTA 1242  
 QY 395 AspIleSerIleGlnValGlyGlnAlaSerIleLeuValProGlnLeuYsMetGlnIle 414  
 DB 1243 GACATATCAAAAGAGTTGGAGAGCTTCCATTAAGTACCAATTAATAATGAGATA 1302

Search completed: November 11, 2005, 03:41:03  
 Job time : 895 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 11, 2005, 03:23:05 ; Search time 5350 Seconds  
(without alignments)  
2945.534 Million cell updates/sec

Title: US-10-649-273-2  
Perfect score: 414  
Sequence: 1 MLILTKTAGVFFPKSKRKY.....DISKEVGASIVPOLKMEI 414

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 segs, 19032134700 residues

Word size: 274

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cgn2.1/USPRO.spool.p/US10649273/runat.02112005.091641.16886/app.query.fasta\_1.583  
-DB=EST -QFMT=fastap -SUFFIX=oligp2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCCALIGN=200 -THR SCORE=quality -THR MIN=274 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273@cgn2.1 1 5180 @runat.02112005.091641.16886 -NCPUs=6 -ICPU=3  
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :  
EST :  
1: gb\_est1 : \*  
2: gb\_est2 : \*  
3: gb\_hic : \*  
4: gb\_est3 : \*  
5: gb\_est4 : \*  
6: gb\_est5 : \*  
7: gb\_est6 : \*  
8: gb\_gss1 : \*  
9: gb\_gss2 : \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
-----					
No matches found					

Search completed: November 11, 2005, 07:15:10  
Job time : 5350 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 11, 2005, 03:22:15 ; Search time 7521 Seconds

(without alignments)  
2667.260 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 414  
Sequence: 1 MLILTKTAGVFFPKSRKRVY.....DISKEVGASINVPOLKMEI 414

Scoring table:

	OLIGO	Xgapop 60.0	Xgapext 60.0
		Xgapop 60.0	Xgapext 60.0
		Fgapop 6.0	Fgapext 7.0
		Delop 6.0	Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Word size: 274

Total number of hits satisfying chosen parameters: 3

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Command line parameters:

```
-MODSL=frame+ p2n.model -DEV=xlp
-Q=/cgn2.1/USPRO.spool_p/US10649273/runat.02112005.091640.16872/app_query.fasta_1.583
-DB=genEmbl -QFMT=fastap -SUFFIX=oligp2n.rge -MINMATCH=0.1 -LOOPCL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=274 -ALIGN=15 -MODE=LOCAL
-OUTMT=ptlo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10649273 @CGN 1.1.560 @runat.02112005.091640.16872 -NCPU=6 -ICPU=3
-NO_MMAP -LRGEOQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC
-DEV_TIMEOUT=150 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7
```

Database :

```
GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_ets:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	414	100.0	BC011904	Homo sapi
2	414	100.0	AR428803	Sequence
3	340	82.1	AR541929	Sequence

#### ALIGNMENTS

RESULT 1  
LOCUS BC011904  
DEFINITION Homo sapiens O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA clone MGC:20293 IMAGE:4121450), complete cds.  
ACCESSION BC011904  
VERSION BC011904.2 GI:40225818  
KEYWORDS MGC.  
SOURCE Homo sapiens (human).  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

#### REFERENCE

1 (bases 1 to 1908)  
Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heleth, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Udell, T.B., Toshiyuki, S., Carninci, P., Prange, C., Kaha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McKwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Ketterman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywnicki, M.I., Skalska, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.U. and Marra, M.A., Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

#### TITLE

JOURNAL PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

#### REMARK

COMMENT  
Contact: MGC help desk  
Email: cgabs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: http://www.nisc.nih.gov/  
Contact: nisc\_mgc@nigr.nih.gov  
Ahter, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Khong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeri, B., Macriani, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantrop, S., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

NIH-MGC Project URL: http://mgc.nci.nih.gov  
On Dec 19, 2003 this sequence version replaced gi:15080281.  
Contact: MGC help desk  
Email: cgabs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: http://www.nisc.nih.gov/  
Contact: nisc\_mgc@nigr.nih.gov  
Ahter, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Khong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeri, B., Macriani, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantrop, S., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

#### FEATURES

source  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov  
Series: IRAL Plate: 28 Row: 3 Column: 22.  
Location/Qualifiers  
1..1908  
/organism="Homo sapiens"  
/mol\_type="mRNA"

```

      /db_xref="taxon:9606"
      /clone="MGC:20293 IMAGE:4121450"
      /tissue_type="Muscle, rhadomyosarcoma"
      /clone_id="N1H_MGC_17"
      /lab_host="DH10B-R"
      /note="Vector: pOTB7"
      1..1908
      /gene="OSGEP1"
      /db_xref="locusID:64172"
      104..1148
      /gene="OSGEP1"
      /codon_start=1
      /product="OSGEP1 protein"
      /protein_id="AAH1904.1"
      /db_xref="GI:15080282"
      /db_xref="locusID:64172"
      /translation="MLILTKTAGVFFKSKKVEPLNSFNHPTGLLHKIVLGET
      SCDDPAADVDETVGEALHSOTEVHLKGVIVPAQOQHRENIQRIYOEALAS
      GVSFSDLSAIAITIKPGLASGLVQLQKPFIPIHMEALITRLTNK
      VEPRLVLLISGHCILALVGVSDFLIGSLDAPGDMLKVARSLIKHPEST
      MSGKALSHLAKQGRFRPDKPRLHAKNCDSPTGLOHTYDKIMKEKEGIEG
      QILSSADIAAVQHTMACHLVKRTHRALPCKQDLPONNAVLYASGVASNFIR
      RALEITNATQCTLLCPPLCTDNGIMIAMNGIERLRAGLGLHDIEGIREPKPL
      GVDISKEVGEASIKVPOLKMEI"
      215..1111
      /gene="OSGEP1"
      /note="Peptidase_M22, Region: Glycoprotease family"
      /db_xref="CDD:pfam00814"

misc_feature
      /gene="OSGEP1"
      /note="Peptidase_M22, Region: Glycoprotease family"
      /db_xref="CDD:pfam00814"

Alignment Scores:
Pred. No.:      0      Length:      1908
Score:          414.00      Matches:      414
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              9      Gaps:      0

US-10-649-273-2 (1-414) x BC011904 (1-1908)
QY      1 MetLeuileLeuThrThyThraIaGIValPhePheLySProSeLySArgLySValTyr 20
DB      104 ATGCTAATCTTACCTAAGCTCAGAGAGTTTAAAAACATCAAGAAAGAAAGTTAT 163
QY      21 GIuPheLearSerePheSnpheHisProGIYThrLeuPheLeuHisLySLeValLeu 40
DB      164 GAATTTTAAAGATTATTTTCTCTGGAACACTATTCTCATATAAATAAGATTG 223
QY      41 GIYlIeGIuThrSerCySaSpaSpThraIaIaIaValVaIaSpGIuThrGIYaSpVal 60
DB      224 GGAATTGAAACTAGTGTATGATACAGACGCTGCTGTGTGATGAATACTGGAATGTG 283
QY      61 LeuGIuGIuAlaIleHisSerGIuThrGIuValHisLeuLySThrGIYlIeValPro 80
DB      284 TTGGGAGAAAGCAATACATTCCTCAAACTGAAGTTCAATTTAAAAACAGGTGGATTGTCCT 343
QY      81 ProAlaIaGIuGIuLeuHisSargGIuAsnIleGIuArgIleValGIuGIuAlaLeuSer 100
DB      344 CCAGACGCTCAACAGCTTCACAGAGAAATATTCAAGAAATAGTACAAAGAGCTCTTCT 403
QY      101 AlaSerGIYValSerProSeRAsPLeuSerAlaIleAlaThrThrIleLySProGIYLeu 120
DB      404 GCACAGTGAAGTCTCTTCAAGTGCCTCTCAGCAATTGCACTCCATPAAAAACAGGACTT 463
QY      121 AlaLeuSerLeuGIYValGIuLeuSerPheSerLeuGIuLeuValGIYGIuLeuLyS 140
DB      464 GCTTTAAGCTTGAGAGTGAGCTTTATCATTTTACCTTACAGCTGTGGACGATTAAAAAG 523
QY      141 ProPheIleProIleHisIsmetGIuAlaHisAlaLeuThrIleArgLeuThrAsnLyS 160
DB      524 CCATTCATTCCTCATTCATATAGAGGCTCATGCACTTCTATTAGGTTGACCAATPAA 583
QY      161 ValGIuPheProPheLeuValLeuLeuIleSerGIYGIYHisCySLeuLeuAlaLeuVal 180
```

```

DB      584 GTAGATTTCCCTTTTATTTCTTTGATTTTCGAGGCTCAGTGTGTCGATTAGT 643
QY      181 GIuGIYValSerAsPheLeuLeuGIYLySLeuSerLeuAspIleAlaProGIYAsPmet 200
DB      644 CAAGAGTTTCAGATTTTCTGCTTTTGAAAGCTCTTTGGACATACACAGGTGACATG 703
QY      201 LeuAsPlyValAlaArgArgLeuSerLeuIleLySHisProGIYCySLeuThrMetSer 220
DB      704 CTTGACAAAGTGGCAAGAAAGACTTTCTTTAATPAAACATCCAGAGTCTCACATGAGT 763
QY      221 GIYGIYLySAlaIleGIuHisLeuAlaLySLeuGIYAsnArgPheHisPheAspIleLyS 240
DB      764 GGTGGGAAAGCCATPAAACATTTGGCCAAACAGAGAAATAGATTCAATTTGACATCAA 823
QY      241 ProPheLeuHisHisAlaLySAsnCySaSPheSerPheThrGIYLeuGIuHisValThr 260
DB      824 CTTCCCTTCATCATCTPAAATATGTATTTCTTTTACTGACTTCACACGTTACT 883
QY      261 AsPlyIleIleMetLySlySGIYLySGIuGIYlIeGIYLySGIYGIuIleLeuSer 280
DB      884 GATTAATTAATATGAAAAAGAAAAAGAAAGATTTGAGAAAGGGCAATCTGTCT 943
QY      281 SerAlaIaAspIleAlaAlaThrValGIuHisThrMetAlaCySHisLeuValLySArg 300
DB      944 TCAGACAGACAGACATTCCTGCCACAGTACAGACACATGCGCATGTCATTGTGAAGA 1003
QY      301 ThrHisArgAlaIleLeuPheCySlySGIuArgAsPLeuLeuProGIuAsnAlaVal 320
DB      1004 ACACATCGGCTATTTCTGTTTGTAAAGACAGAGACTTGTACTTCAATTAATAGCATA 1063
QY      321 LeuValAlaSerGIYGIYValAlaSerAsPheTyrIleArgArgAlaLeuGIuIleLeu 340
DB      1064 CTGCTTCATCTGCTGTGTGCGAAGTATTTCTATTCGACAGAGCTCTGGAATTTTA 1123
QY      341 ThrAsnAlaThrGIYThrLeuLeuCySProProArgLeuCySThrAspArgLyS 360
DB      1124 ACAAAGCAACACAGTGCATTTGTTGTGTCTCTCCACAGACTATGCATGAATAGGC 1183
QY      361 IleMetIleAlaThrPheArgLyIleGIuArgLeuArgAlaLyLeuGIYlIeLeuHisAsp 380
DB      1184 ATTATGATTCATGATGATGATGATTAATGAAGCTACGTGCTGTGGCATTTTATCATGAC 1243
QY      381 IleGIuGIYlIeArgTyrGIuProLyCySProLeuGIYValAsPLeuSerGIYVal 400
DB      1244 ATGAAAGGCAATCCGCTATGAAACAAATGTCTCTTGAATGACATATCAAAAGAGTT 1303
QY      401 GIYGIuAlaSerIleLySValProGIuLeuLySmetGIuIle 414
DB      1304 GAGAAAGCTTCATPAAAGATACACACATTTAAAAATGAGATA 1345

RESULT 2
AR428803      2197 bp      DNA      linear      PAT 18-DEC-2003
LOCUS      AR428803
DEFINITION      Sequence 1 from patent US 6642041.
ACCESSION      AR428803
VERSION      AR428803.1 GI:40186589
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 2197)
AUTHORS      Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.
TITLE      Polynucleotides encoding a novel metalloprotease, Mp-1
JOURNAL      Patent: US 6642041-A 1 04-NOV-2003;
FEATURES
source      1..2197
            /organism="unknown"
            /mol_type="genomic DNA"

ORIGIN
Alignment Scores:      0      Length:      2197
Pred. No.:      0
```

Score: 414.00 Matches: 414  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-10-649-273-2 (1-414) x AR428803 (1-2197)

QY	1	MettellileuThrlThraAglYValPhePheLysProSerLysArgLysValTyr	20
DB	231	ATGCTAACTTGACAGACAGCTGAGAGCTTTTAAACCATCAAAAAGAAAGTTAT	290
QY	21	GIUPheLysSerPheAspPheHisProGlyThrLeuPheLysHisLysIleValLeu	40
DB	291	GAAATTTTAAAGAGTTTAAATTTTCACTGGAACCTATTTCTTATAAAATAGATG	350
QY	41	GIYIleGluThrSerCysAspAspThrAlaAlaValAspGluThrGlyAsnVal	60
DB	351	GGAATTTGAACTAGTTGTGATGATACAGACAGCTGCTGTGTGATGAACTGGAAATGTG	410
QY	61	LeuGIYIleGluAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyIleValPro	80
DB	411	TTGGAGAGAGCAATATCATTTCCAACTGAACTTATTAACAGGTGGATTTGTTCT	470
QY	81	ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer	100
DB	471	CCACAGCTCAACAGCTTCAAGAGAAATATTCAACGAATGATCAAGAGCTCTTTCT	530
QY	101	AlaSerGlyValSerProSerAspSerLeuSerAlaIleAlaThrThrIleLysProGlyLeu	120
DB	531	GCCAGTGAAGTCTTCCAAAGTACCTCTCAGCAATTCACATCAATAAACAGAGACTT	590
QY	121	AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys	140
DB	591	GCTTTAAGCTGGAGAGTGGCTTATCATTTACCTTACAGCTGTGAGACAGTTAAAG	650
QY	141	ProPheIleProIleHisIleIleGluAlaHisAlaLeuThrIleArgLeuThrAsnLys	160
DB	651	CCATTCAATCCCATTCATCATATGAGGCTCATGACATTACTATTAAGTTGACCAATAA	710
QY	161	ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal	180
DB	711	GTAAGATTTCTTTTAACTTTTGAATTTCTGAGGCTCAGCTGTGTGGCATTAAGT	770
QY	181	GlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet	200
DB	771	CAAGAGATTCAAGATTTCTGCTTCTTGAAAGCTTTTGAGACATACAGCACAGTGAAGT	830
QY	201	LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThrMetSer	220
DB	831	CTTGACAAAGTGGCAGAAAGACTTTCTTTAATTAACATCCAGAGTGTCCACATAGAT	890
QY	221	GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys	240
DB	891	GGTGGAAAGCCATAGAACATTTGGCCAAACAGAGAAATAGATTTCAATTTGACATCAA	950
QY	241	ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr	260
DB	951	CTCTCCCTGACATCAAGCTAAATATTTGATTTCTTTTACCTGACCTTCAACACGTACT	1010
QY	261	AspLysIleIleMetLysLysGlyLysGlyGlyIleGlyLysGlyGlnIleLeuSer	280
DB	1011	GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1070
QY	281	SerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArg	300
DB	1071	TCAGACACAGCATTTGCTGCACAGTACAGCACACAAATGCGATGATCTTGTGAAAAG	1130
QY	301	ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAsnAlaVal	320
DB	1131	ACACATGGGGCTATTTCTTTTGTAAAGCAGAGACTGTGTATCTCAAAATATATCAGTA	1190
QY	321	LeuValAlaSerGlyValAlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeu	340

DB	1191	CTGATGCACTGAGTGTGTCGCAATTAATTCGACAGCTCTGGAATTTTA	1250
QY	341	ThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaArgLeuCysThrAspAsnGly	360
DB	1251	ACAAAGCAACACAGTGCATTTGTTGTGTCTCTCCAGACATATGACATGATATGCC	1310
QY	361	IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp	380
DB	1311	ATTATGATTCAGATGAAATGATTAATTAAGACTAGTCTGCTGGCATTTTACATGAC	1370
QY	381	IleGluGlyIleArgArgGlyProLysCysProLeuGlyValAspIleSerLysGlyVal	400
DB	1371	ATGAAAGCATCCGCTATGAAACCAAAATGTCTCTTGAGTGAACATATCAAAAGAGTT	1430
QY	401	GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle	414
DB	1431	GGAAGACTTCCATTAATAAGTACCAATTAATAATGAGATTA	1472

RESULT 3  
 AR541929 1416 bp DNA linear PAT 08-OCT-2004  
 LOCUS Sequence 177 from patent US 6743619.  
 DEFINITION AR541929  
 ACCESSION AR541929  
 VERSION AR541929.1 GI:53934009  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1416)  
 AUTHORS Tang,Y.T., Zhou,P., Goodrich,R., Liu,C., Asundi,V., Ren,F., Zhang,J., Zhao,Q.A., Yang,Y., Xue,A.J., Wehrman,T., Wang,J.-R., Wang,D. and Drmanac,R.T.  
 TITLE Nucleic acids and polypeptides  
 JOURNAL Patent: US 6743619-A 177 01-JUN-2004;  
 FEATURES  
 source location/Qualifiers  
 1..1416  
 /organism="unknown"  
 /mol\_type="genomic DNA"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 1416  
 Score: 340.00 Matches: 340  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 82.13% Indels: 0  
 DB: 6 Gaps: 0

QY	75	ThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIle	94
DB	283	ACAGGGGAGATTTCTTCTCAGAGCTCAACAGCTTCAAGAGAAATATTAACGAATTA	342
QY	95	ValGlnGluAlaLeuSerAlaSerLysValSerProSerAspLeuSerAlaIleAlaThr	114
DB	343	GTAACAAGAGCTTTCTTCTCAGAGAGTCTTCAAGTAACTCTCAAGAAATGCAACT	402
QY	115	ThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeu	134
DB	403	ACCATTAACACAGACTTCTTAAAGCTGAGAGTGGCTTATCATTTAGCTTACAGCTG	462
QY	135	ValGlyGlnLeuLysLysProPheIleProIleHisIleMetGluAlaHisAlaLeuThr	154
DB	463	GTAGACAGATTAATAAAAGCATTCATTCATTCATCATATGAGAGCTCATGCACTTAAT	522
QY	155	IleArgLeuThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyHis	174
DB	523	ATTAGATGACCAATTAATAAGTAAATTTCTTTTGTGATTTCTTGTGAGAGTCAAC	582
QY	175	CysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAsp	194

```
Db      583 TGTCGTGGCATTAGTTCAAGAGATTTCAGATTTTCGCTTGTGAAAGCTTTGAGC 642
Qy      195 ILeAlaProGlyAspMetLeuAspLysValAlaArgLeuSerLeuIleLysHisPro 214
Db      643 ATAGCACCAAGTGACATGCTTGACAGAGGTGACAGAGAGCTTTTAATAAACAATCCA 702
Qy      215 GluCysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArg 234
Db      703 GAGTGCTCCACCAAGAGTGAGTGAGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGA 762
Qy      235 PheHisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThr 254
Db      763 TTTCAATTTTGACATCAAACTCCCTTGCAATGCTTAATAAAATTGCAATTTTCTTTTACT 822
Qy      255 GlyLeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluLysGluIleGlu 274
Db      823 GGACTTCAAACAGTTACTGATTAATAATTAATGAAGAAAGAGAGAGATTTGAG 882
Qy      275 LysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAla 294
Db      883 AAGGGCAAAATCGTCTTCAGCAGACAGCATTTGCTGCCACAGTACAGCACACATGGCA 942
Qy      295 CysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeu 314
Db      943 TGTCACTCTGTGAAGAAGACACATCGGGCTATTCTGTTTGTAGCAGAGAGCTTGTTA 1002
Qy      315 ProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArg 334
Db      1003 CCTCAAAATAATGACGTACTGCTGTCATCTGGTGGTCGCAAGTAACTTATATATCCGC 1062
Qy      335 ArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArg 354
Db      1063 AGAGCTCTGGAATTTTAACAACACGCAACAGTGCATTTGTGTCTCTCTCCACAG 1122
Qy      355 LeuCysThrAspAsnGlyIleMetIleAlaTyrPasnGlyIleGluArgLeuArgAlaGly 374
Db      1123 CTATCCACTGATATGGCATTTGATGCAATGAAATGATATGAAGAATACTACCTGCTGGC 1182
Qy      375 LeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyVal 394
Db      1183 TTGGGCATTTTATCATGACATAGAGGATCCGCTATGAACCAAAATGTCTCTTGGAGTA 1242
Qy      395 AspIleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
Db      1243 GACATATCAAAAGAGTGGAGAAAGCTTCATATAAAGTACCAATTAATAAATGAGAGATA 1302
```

Search completed: November 11, 2005, 05:46:33  
Job time : 7527 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:02:11 ; Search time 71.9739 Seconds

(without alignments)  
1434.756 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414

Perfect score: 1385  
Sequence: 1 MEAHALTRLTNRYKPPFLV.....DISKEVGASIKVQLKMEI 267

Scoring table: BLASTN62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_16Dec04: \*  
1: geneseqp1980s: \*  
2: geneseqp1990s: \*  
3: geneseqp2000s: \*  
4: geneseqp2001s: \*  
5: geneseqp2002s: \*  
6: geneseqp2003as: \*  
7: geneseqp2003bs: \*  
8: geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1385	100.0	267	5	ABG96487 Novel hum
2	1385	100.0	414	5	ABG96478 Novel hum
3	1385	100.0	414	5	ABR05481 Human O-s
4	1385	100.0	414	6	ABJ26654 Human pro
5	1348	97.3	414	5	AAE29234 Human gly
6	1348	97.3	414	6	ABG71161 Human gly
7	1348	97.3	414	6	ABU09569 Human pro
8	1105	79.8	364	6	ADA54471 Human pro
9	415.5	30.0	409	4	ABB69133 Drosophila
10	409.5	29.6	463	3	AAV52216 Arabidops
11	398.5	28.8	245	3	AAV52216 Arabidops
12	398.5	28.8	439	3	AAV52216 Arabidops
13	398.5	28.8	444	3	AAV52216 Arabidops
14	398.5	28.8	444	3	AAV52216 Arabidops
15	398.5	28.8	444	3	AAV52216 Arabidops
16	398.5	28.8	444	3	AAV52216 Arabidops
17	398.5	28.8	444	3	AAV52216 Arabidops
18	398.5	28.8	444	3	AAV52216 Arabidops
19	398.5	28.8	444	3	AAV52216 Arabidops
20	398.5	28.8	444	3	AAV52216 Arabidops
21	398.5	28.8	444	3	AAV52216 Arabidops
22	398.5	28.8	444	3	AAV52216 Arabidops
23	398.5	28.8	444	3	AAV52216 Arabidops
24	398.5	28.8	444	3	AAV52216 Arabidops
25	398.5	28.8	444	3	AAV52216 Arabidops

26	289.5	20.9	342	4	AAU35450 Haemophil
27	289.5	20.9	342	6	ABU30280 Protein e
28	289.5	20.9	340	6	ABU40514 Protein e
29	289.5	20.3	357	7	ADF06228 Bacterial
30	279.5	20.2	337	4	AAU38187 Salmonell
31	279.5	20.2	337	6	ABU47551 Protein e
32	279.5	20.2	337	6	ABU50237 Protein e
33	279.5	20.0	335	6	ABU27480 Protein e
34	277	20.0	309	6	ABU33614 Protein e
35	277	20.0	341	6	ABU40069 Protein e
36	276	19.9	421	5	ABG96491 Novel hum
37	275.5	19.9	337	4	AAU34711 E. coli c
38	275.5	19.9	337	6	ABU28771 Protein e
39	270.5	19.5	325	2	AAE26325 Glycoprot
40	270.5	19.5	325	3	AAV52203 Pasteurel
41	270.5	19.5	337	3	AAV52204 Bacterioph
42	268.5	19.4	343	7	ABO62704 Klebsiell
43	266.5	19.2	338	6	ABM67812 Phototrab
44	263.5	19.0	354	6	ABU37844 Protein e
45	263	19.0	341	6	ABU41689 Protein e

#### ALIGNMENTS

RESULT 1	ABG96487	strand: protein; 267 AA.
ID	ABG96487	
AC	ABG96487	
DT	11-DEC-2002	(first entry)
DB	Novel human metalloprotease MP1 fragment #1.	
XX	Metalloprotease, MP-1; immune disorder; glutamate transport; cancer;	
XX	motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;	
XX	reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;	
XX	genital wart; metabolic disorder; premature puberty; Kallman syndrome;	
XX	Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;	
XX	Parinson's disease; Huntington's disease; Tourette syndrome; sepsis;	
XX	liver disease; renal disease; immune disorder; rheumatoid arthritis;	
XX	acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;	
XX	emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;	
XX	neurological disorder.	
OS	Homo sapiens.	
PN	W0200272751-A2.	
PD	19-SEP-2002.	
PF	05-FEB-2002; 2002WO-US003353.	
PR	05-FEB-2001; 2001US-0266518P.	
PR	10-APR-2001; 2001US-0282814P.	
PI	(BRIM) BRISTOL-MYERS SQUIBB CO.	
PI	Chen J, Feder J, Nelson TC, Duclos F, Krystek S;	
DR	MP1: 2002-723329/78.	
DR	N-PSDB; ABS76639.	
PT	New isolated nucleic acid encoding MP-1 protein, useful for preventing,	
PT	treating, or ameliorating diseases associated with aberrant	
PT	metalloprotease activity, e.g. immune, metabolic, inflammatory and	
PT	neurological disorders.	
PS	Claim 5; Page 29; 473pp; English.	
CC	The invention describes an isolated nucleic acid molecule (I) encoding a	
CC	metalloprotease (MP-1). (I) is useful for preventing, treating, or	

CC ameliorating a medical condition, particularly an immune disorder, an  
CC aberrant glutamate transport or motor neuron disorder, such as  
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
CC condition. The compositions and methods are also useful for diagnosing,  
CC prognosticating, treating, ameliorating and/or treating disorders  
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
CC sequence of a metalloprotease MPI protein  
XX  
SQ Sequence 267 AA;  
  
Query Match 100.0%; Score 1385; DB 5; Length 267;  
Best Local Similarity 100.0%; Pred. No. 1.4e-147;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MEAHALTRLTNKNKPEPFLVLLISGGHCLALVQVSDPFLILGKSLDIPAGMDLVARR 60  
DB 1 MEAHALTRLTNKNKPEPFLVLLISGGHCLALVQVSDPFLILGKSLDIPAGMDLVARR 60  
QY 61 LSLIKPECSMTSGGKAI EHLAKQGNRFPDIKPLHAKNCDPSFTGLQHTVDKTIIMKK 120  
DB 61 LSLIKPECSMTSGGKAI EHLAKQGNRFPDIKPLHAKNCDPSFTGLQHTVDKTIIMKK 120  
QY 121 EKEBGEIKGQILSSADIAATVQHTWACHLVKTRTHAILFCQRDILLPQNNAVLVASGV 180  
DB 121 EKEBGEIKGQILSSADIAATVQHTWACHLVKTRTHAILFCQRDILLPQNNAVLVASGV 180  
QY 181 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLGIHDI EGRYE 240  
DB 181 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLGIHDI EGRYE 240  
QY 241 PKCPLGVDSKEVGEASIKVPOLKMEI 267  
DB 241 PKCPLGVDSKEVGEASIKVPOLKMEI 267  
  
RESULT 2  
ABG96478  
ID ABG96478 standard; protein; 414 AA.  
XX  
AC ABG96478;  
XX  
DT 11-DEC-2002 (first entry)  
XX  
DE Novel human metalloprotease MPI.  
XX  
XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
KW neurological disorder.  
XX  
XX Homo sapiens.  
OS  
PN MO200272751-A2.  
XX  
PD 19-SEP-2002.  
XX  
XX 05-FEB-2002; 2002WO-US003353.  
PF  
XX 05-FEB-2001; 2001US-026518P.

PR 10-APR-2001; 2001US-0282814P.  
XX  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
PA  
XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S,  
PI  
XX MPI, 2002-723329/78.  
DR  
XX N-PSDB; ABS76635.  
XX  
PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
PT treating, or ameliorating diseases associated with aberrant  
PT metalloprotease activity, e.g. immune, metabolic, inflammatory and  
PT neurological disorders.  
XX  
PS Claim 5; Fig 1A-C; 473bp; English.  
XX  
CC The invention describes an isolated nucleic acid molecule (I) encoding a  
CC metalloprotease (MP-1). (I) is useful for preventing, treating, or  
CC ameliorating a medical condition, particularly an immune disorder, an  
CC aberrant glutamate transport or motor neuron disorder, such as  
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
CC condition. The compositions and methods are also useful for diagnosing,  
CC prognosticating, treating, ameliorating and/or treating disorders  
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
CC sequence of a metalloprotease MPI protein  
XX  
SQ Sequence 414 AA;  
  
Query Match 100.0%; Score 1385; DB 5; Length 414;  
Best Local Similarity 100.0%; Pred. No. 2.8e-147;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MEAHALTRLTNKNKPEPFLVLLISGGHCLALVQVSDPFLILGKSLDIPAGMDLVARR 60  
DB 148 MEAHALTRLTNKNKPEPFLVLLISGGHCLALVQVSDPFLILGKSLDIPAGMDLVARR 207  
QY 61 LSLIKPECSMTSGGKAI EHLAKQGNRFPDIKPLHAKNCDPSFTGLQHTVDKTIIMKK 120  
DB 208 LSLIKPECSMTSGGKAI EHLAKQGNRFPDIKPLHAKNCDPSFTGLQHTVDKTIIMKK 267  
QY 121 EKEBGEIKGQILSSADIAATVQHTWACHLVKTRTHAILFCQRDILLPQNNAVLVASGV 180  
DB 268 EKEBGEIKGQILSSADIAATVQHTWACHLVKTRTHAILFCQRDILLPQNNAVLVASGV 327  
QY 181 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLGIHDI EGRYE 240  
DB 328 ASNFYIRRALIELTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLGIHDI EGRYE 387  
QY 241 PKCPLGVDSKEVGEASIKVPOLKMEI 267  
DB 388 PKCPLGVDSKEVGEASIKVPOLKMEI 414  
  
RESULT 3  
ABB05481  
ID ABB05481 standard; protein; 414 AA.  
XX  
XX ABB05481;  
XX  
DT 19-APR-2002 (first entry)  
XX  
XX Human O-sialoglycoproteinase-like protein SEQ ID NO.2.  
DE  
XX Human; O-sialoglycoproteinase-like protein; OSGPUP; enzyme.  
XX



OS Homo sapiens.  
XX CN1318550-A.  
XX  
XX 24-OCT-2001.  
XX  
XX 19-APR-2000; 2000CN-00106834.  
XX  
XX 19-APR-2000; 2000CN-00106834.  
XX  
XX (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.  
XX  
XX Mao Y, Xie Y;  
XX  
XX WPI; 2002-115090/16.  
XX  
XX N-PSDB; ABA93268.  
XX  
XX O-sialoglycoproteinase-like protein and encoding polynucleotide, useful  
XX for diagnosing, preventing and treating related diseases.  
XX  
XX Claim 1; Page 30-31 (Disclosure); 38pp; Chinese.  
XX  
XX The present sequence represents human O-sialoglycoproteinase-like protein  
XX (OSGPLP). The present invention also describes: (1) the preparation of  
XX the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the  
XX prevention and/or treatment of related diseases; (4) utilizing the OSGPLP  
XX protein in screening its agonist, excitomotor and inhibitor and preparing  
XX an antibody against the OSGPLP protein; and (5) the use of the OSGPLP  
XX polynucleotide sequences, proteins, agonists, excitomotors, inhibitors  
XX and antibodies in treating diseases related to the abnormal OSGPLP gene  
XX and in preparing the medicine composite for the treatment  
XX  
XX Sequence 414 AA;

Query Match 100.0%; Score 1385; DB 5; Length 414;  
Best Local Similarity 100.0%; Pred. No. 2.8e-147;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAHALTITLTKVPPFLVLLISGHCLLAVGVSPFLIGKSLDIARQDMLDKVAR 60  
DB 148 MEAHALTITLTKVPPFLVLLISGHCLLAVGVSPFLIGKSLDIARQDMLDKVAR 207  
QY 61 LSLIKHECSTMSGKAIIEHLAKOGRFHDIKPELHAKNCDSFTGQHQHTDKIMRK 120  
DB 208 LSLIKHECSTMSGKAIIEHLAKOGRFHDIKPELHAKNCDSFTGQHQHTDKIMRK 267  
QY 121 EKEEGIEKGQILSSAADIAATVQHTMACHLVKTRTRAILFCQRDILLPQNNAVLVASGV 180  
DB 268 EKEEGIEKGQILSSAADIAATVQHTMACHLVKTRTRAILFCQRDILLPQNNAVLVASGV 327  
QY 181 ASNYIRRAAEILTNATQCTLLCPPRLCTNGIMANGIRLPAAGLILHDIGIRYE 240  
DB 328 ASNYIRRAAEILTNATQCTLLCPPRLCTNGIMANGIRLPAAGLILHDIGIRYE 387  
QY 241 PKCPGLVDISKVGEVASIKVPOLKMEI 267  
DB 388 PKCPGLVDISKVGEVASIKVPOLKMEI 414

RESULT 4  
ABJ26654  
ID ABJ26654 standard; protein; 414 AA.  
XX  
XX AC ABJ26654;  
XX  
XX DT 01-MAY-2003 (first entry)  
XX  
XX DE Human protein modification + maintenance molecule protein SEQ ID No 8.  
XX  
XX CYTOSTATIC; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;  
XX cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;  
XX antitumor; hepatotropic; gynecological; antibacterial; virucide;  
XX protozoacide; antiparasitic; cell proliferative disease; PMOD;

KW protein modification and maintenance molecule; immunogenic fragment;  
KW cancer; autoimmune; inflammatory disease; neurological disorder;  
KW gastrointestinal; developmental; vesicle trafficking disorder; infection;  
KW protein-protein interaction; drug-target interaction;  
KW gene expression profile; human.

OS Homo sapiens.  
XX  
XX WO200300844-A2.  
XX  
XX 03-JAN-2003.  
XX  
XX 18-JUN-2002; 2002WO-US019360.  
XX  
XX 22-JUN-2001; 2001US-0300508P.  
XX 06-JUL-2001; 2001US-0303445P.  
XX 13-JUL-2001; 2001US-0305405P.  
XX 09-AUG-2001; 2001US-0311442P.  
XX 24-AUG-2001; 2001US-0314821P.  
XX 29-AUG-2001; 2001US-0315992P.  
XX 03-MAY-2002; 2002US-0378205P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Gandhi AR, Kable AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BW,  
XX Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DAM, Lee EA, Yue H;  
XX Foreythe JF, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J;  
XX Thangavelu K, Gietzen KJ, Ding L, Baughn MR, Borowaky ML, Yao MG;  
XX Walla NK, Mason PM, Guturjan R, Lee S, Becha SD, Lee ST, Tran UK,  
XX Elliott VS, Luo W, Sprague WW, Tang YT, Lu Y, Zebardjian Y;  
XX  
XX WPI; 2003-184039/18.  
XX  
XX N-PSDB; ABR23207.  
XX  
XX New isolated human PMOD polypeptide and polynucleotide, useful for  
XX diagnosing, treating and preventing diseases or conditions associated  
XX with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and  
XX infections.  
XX  
XX Claim 63; Page 182-183; 225pp; English.

XX The invention relates to an isolated polypeptide comprising: any of 28  
XX sequences of 48-1256 amino acids; a natural amino acid sequence at least  
XX 90% identical to the 28 amino acid sequences, 94% identical to a sequence  
XX of 703 or 267 amino acids, 96% identical to a sequence of 414 amino  
XX acids, or 97% identical to a sequence of 242 amino acids, all given in  
XX the specification; or a biologically active or immunogenic fragment of  
XX the isolated polypeptide. The polypeptides and polynucleotides are useful  
XX in diagnosing, treating and preventing diseases or conditions associated  
XX with the decreased expression of protein modification and maintenance  
XX molecules (PMOD), such as cell proliferative diseases (e.g. cancer,  
XX atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,  
XX allergies), neurological disorders (e.g. stroke, Parkinson's disease,  
XX epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.  
XX endometriosis), developmental, vesicle trafficking disorders, and  
XX infections (e.g. bacterial, viral, parasitic, protozoal). These are also  
XX useful in assessing the effects of exogenous compounds on the expression  
XX of nucleic acid and amino acid sequences of PMOD. The PMOD or its  
XX fragments are useful in screening compounds for effectiveness as agonist  
XX or antagonist of the polypeptides, or in altering the expression of the  
XX target polynucleotide and compounds that specifically bind to or modulate  
XX the activity of the polypeptide. The microarray is useful in monitoring  
XX or measuring protein-protein interactions, drug-target interactions, and  
XX gene expression profiles. This sequence represents a human PMOD protein  
XX of the invention  
XX

SO Sequence 414 AA;

Query Match 100.0%; Score 1385; DB 5; Length 414;  
Best Local Similarity 100.0%; Pred. No. 2.8e-147;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MEAHALTITLTKVPPFLVLLISGHCLLAVGVSPFLIGKSLDIARQDMLDKVAR 60

Db 148 MEAHALTRLTNKVEPFLVLLISGGHCLTALVQGVSDPFLLGKSLDIAPGMDLVARR 207  
Qy 61 LSLIKHEPCSTMSGKAIEHLAKOGRFHPDIKPLHAKNCDPFSFTGLOHTYDKTIMKK 120  
Db 208 LSLIKHEPCSTMSGKAIEHLAKOGRFHPDIKPLHAKNCDPFSFTGLOHTYDKTIMKK 267  
Qy 121 EKEGIEKGQILSSAADIAATVQHTMACHLVKTRTHRAILFCQKRDLLPQNNAVLVASGV 180  
Db 268 EKEGIEKGQILSSAADIAATVQHTMACHLVKTRTHRAILFCQKRDLLPQNNAVLVASGV 327  
Qy 181 ASNFYIRRALLETITNATQCTLLCPPLRCTDNGMIANNGIERLRAGLGIHDIIGIRYE 240  
Db 328 ASNFYIRRALLETITNATQCTLLCPPLRCTDNGMIANNGIERLRAGLGIHDIIGIRYE 387  
Qy 241 PKCPLGVDISKVEGASIKVPOLKMEI 267  
Db 388 PKCPLGVDISKVEGASIKVPOLKMEI 414

## RESULT 5

AAE29234 standard; protein; 414 AA.

AAE29234;

27-JAN-2003 (first entry)

Human glycoprotease 28472 protein.

Human; adenosine deaminase; seven transmembrane domain receptor; cancer;  
7TM; glycoprotease; immune disorder; IGA deficiency; allergy; arylthymia;  
rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;  
hyperendion; ischaemic heart disease; obesity; myocardial infarction;  
endothelial cell disorder; Grave's disease; psoriasis; brain disorder;  
Parkinson's disease; Alzheimer's disease; haematopoietic disorder;  
cerebral oedema; metabolic disorder; liver disorder; platelet disorder;  
chromosome mapping; tissue typing; gene therapy; neuroprotective;  
cytostatic; anorectic; cardiant; haemostatic.

Homo sapiens.

Location/Qualifiers

1..108 /note= "Non-transmembrane domain; N-terminal cytoplasmic domain"

38..369 /note= "Endopeptidase O-sialoglycoprotein hydrolase metalloprotease zinc glycoprotease sialoglycoprotease domain"

109..132 /note= "Transmembrane domain"

133..164 /note= "Non-transmembrane domain; non-cytoplasmic loop"

138..152 /note= "Glycoprotease domain"

165..189 /note= "Transmembrane domain"

190..216 /note= "Non-transmembrane domain; cytoplasmic domain"

317..333 /note= "Transmembrane domain"

334..414 /note= "Non-transmembrane domain"

374..414 /note= "Sialoglycoprotease type domain"

Domain

W0200274960-A2.

26-SEP-2002.

08-NOV-2001; 2001WO-US051427.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PR 08-NOV-2000; 2000US-0246768P.  
PR 08-NOV-2000; 2000US-0246772P.  
PR 15-NOV-2000; 2000US-0249185P.

(MILL-) MILLENNIUM PHARM INC.

Leiby KR, Kapeller-Libermann R, Glucksmann M,

WPI; 2002-759898/82.

N-PSDB; AAD46856.

New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,

useful for diagnosing and treating cancer, immune, cardiovascular,

hematopoietic, brain, pain, metabolic, liver or platelet disorders, and

in pharmacogenomics.

Claim 1; Fig 8; 178bp; English.

The present invention relates to novel 38650, 28472, 5495, 65507, 81588

or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-

protease or seven transmembrane domain (7TM) receptor family members.

Sequences of the invention are useful in diagnosing and treating cancer

or aberrant cellular proliferation and/or differentiation (e.g. colon or

lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid

arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,

hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,

myocardial infarction, thrombus) including endothelial cell disorders

(e.g. psoriasis, Grave's disease), haematopoietic disorders, brain

disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),

pain and metabolic disorders (e.g. obesity), liver disorders or platelet

disorders. They are also useful in screening assays, predictive medicine

(e.g. diagnostic assays, prognostic assays, monitoring clinical trials

and pharmacogenetics) and prophylactic and therapeutic methods. The

nucleic acids may also be used in chromosome mapping, tissue typing and

forensic biology and as surrogate markers. Sequences of the invention are

also used in gene therapy. The present sequence is human glycoprotease

28472 protein

Sequence 414 AA;

Query Match 97.3%; Score 1348; DB 5; Length 414;

Best Local Similarity 97.4%; Pred. No. 4,3e-143;

Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

1 MEAHALTRLTNKVEPFLVLLISGGHCLTALVQGVSDPFLLGKSLDIAPGMDLVARR 60

148 MEAHALTRLTNKVEPFLVLLISGGHCLTALVQGVSDPFLLGKSLDIAPGMDLVARR 207

61 LSLIKHEPCSTMSGKAIEHLAKOGRFHPDIKPLHAKNCDPFSFTGLOHTYDKTIMKK 120

208 LSLIKHEPCSTMSGKAIEHLAKOGRFHPDIKPLHAKNCDPFSFTGLOHTYDKTIMKK 267

121 EKEGIEKGQILSSAADIAATVQHTMACHLVKTRTHRAILFCQKRDLLPQNNAVLVASGV 180

268 KOEGIEKGQILSSAADIAATVQHTMACHLVKTRTHRAILFCQKRDLLPQNNAVLVASGV 327

181 ASNFYIRRALLETITNATQCTLLCPPLRCTDNGMIANNGIERLRAGLGIHDIIGIRYE 240

328 ASNFYIRRALLETITNATQCTLLCPPLRCTDNGMIANNGIERLRAGLGIHDIIGIRYE 387

241 PKCPLGVDISKVEGASIKVPOLKMEI 267

388 PKCPLGVDISKVEGASIKVPOLKMEI 414

RESULT 6

ABG71161

ID ABG71161 standard; protein; 414 AA.

AC ABG71161;

XX 30-JAN-2003 (first entry)

XX

DE Novel human glycoprotease 28472.  
XX  
XX Cancer; aberrant cell proliferation; aberrant cell differentiation;  
XX breast cancer; ovarian cancer; prostate cancer; colon cancer;  
XX lung cancer; immune disorder; heart disorder; cardiovascular disorder;  
XX endothelial disorder; hemopoietic disorder; blood vessel disorder;  
XX brain disorder; pain; metabolic disorder; liver disorder; diabetes;  
XX platelet disorder; carcinoma; sarcoma; leukaemia; Hodgkin's disease;  
XX autoimmune disorder; hypertension; atherosclerosis; heart failure;  
XX myocardial infarction; ischaemic heart disease; Crohn's disease;  
XX Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;  
XX cerebral ischaemia; peripheral neuropathy; Alzheimer's disease;  
XX Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; enzyme.  
OS Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT Domain 138..152  
FT /label= Glycoprotease\_domain  
XX  
XX MO200277233-A2.  
XX  
XX 03-OCT-2002.  
XX  
XX 08-NOV-2001; 2001WO-US046724.  
XX  
XX 08-NOV-2000; 2000US-0246768P.  
XX 08-NOV-2000; 2000US-0246772P.  
XX 15-NOV-2000; 2000US-0249185P.  
XX  
XX (MIL-) MILLENNIUM PHARM INC.  
XX  
XX Leiby KR, Kapeller-Libermann R, Glucksmann M;  
XX  
XX WPI; 2003-029938/02.  
XX  
XX N-PSDB; A6557020.  
XX  
XX New adenosine deaminase, glycoprotease and seven transmembrane domain  
XX PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,  
XX PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or  
XX PT hypertension.  
XX  
XX  
XX Claim 4; Fig 8A-B; 178pp; English.  
XX  
XX  
XX The invention describes isolated 38650, 28472, 5495, 65507, 81588 and  
XX CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The  
XX CC 38650 nucleic acid molecule comprises a sequence encoding adenosine  
XX CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding  
XX CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise  
XX CC sequences that encode a human seven transmembrane domain (7TM). The  
XX CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide  
XX CC sequences are useful for diagnosing, preventing or treating a subject  
XX CC with or at risk of developing a disorder, e.g. cancer or aberrant  
XX CC cellular proliferation and/or differentiation (e.g. breast, ovarian,  
XX CC prostate, colon or lung cancer), immune disorders, heart disorders,  
XX CC cardiovascular disorders, endothelial disorders, hemopoietic disorders,  
XX CC blood vessel disorders, brain disorders, pain and metabolic disorders,  
XX CC liver disorders or platelet disorders. These disorders include carcinoma,  
XX CC sarcoma, leukaemia, Hodgkin's disease, autoimmune disorders,  
XX CC hypertension, atherosclerosis, heart failure, myocardial infarction,  
XX CC ischaemic heart disease, Crohn's disease, Grave's disease, Kawasaki  
XX CC syndrome, Raynaud's disease, aneurysm, cerebral ischaemia, peripheral  
XX CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,  
XX CC cachexia or diabetes. This is the amino acid sequence of the novel human  
XX CC glycoprotease 28472  
XX  
XX  
XX Sequence 414 AA:  
XX  
XX  
XX Query Match 97.3%; Score 1348; DB 6; Length 414;  
XX Best local Similarity 97.4%; Pred. No. 4.3e-143;  
XX Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
XX  
XX 1 MEAAATLRLTNKVEPFLVLLISGHCLLAVQGVSDPFLLGKSLDIAFGDMLDVARR 60

DB 148 MEAAATLRLTNKVEPFLVLLISGHCLLAVQGVSDPFLLGKSLDIAFGDMLDVARR 207  
QY 61 ISLIRHECSTSGGKAIETHLAKQGNRFHFDIKPPLPHAKNCDPSTGQHVTDKIMKK 120  
DB 208 ISLIRHECSTSGGKAIETHLAKQGNRFHFDIKPPLPHAKNCDPSTGQHVTDKIMKK 267  
QY 121 EKEEGIEKQIISAAADIAATVQHTMACLVKRTTRAILFCQRDILLPNNNAVVASGV 180  
DB 268 KQEEGIEKQIISAAADIAATVQHTMACLVKRTTRAILFCQRDILLPNNNAVVASGV 327  
QY 181 ASNFYIRALLETITMAOTCTLCPPRLCTDNGIMANNGIFRLRGIGLIDIEGIRE 240  
DB 328 ASNFYIRALLETITMAOTCTLCPPRLCTDNGIMANNGIFRLRGIGLIDIEGIRE 387  
QY 241 PKCPGVDSKEVGEASIKVPOLKMEI 267  
DB 388 PKCPGVDSKEVGEASIKVPOLKMEI 414  
RESULT 7  
ABU09569  
ID ABU09569 standard; protein; 414 AA.  
XX  
XX ABU09569;  
XX  
XX 08-JUL-2003 (first entry)  
XX  
XX  
XX Human glycoprotease encoded by cDNA 28472.  
XX  
XX  
XX Human; enzyme; cancer; aberrant cellular proliferation; differentiation;  
XX KW immune disorders; heart disorder; brain disorder;  
XX KW cardiovascular disorder; endothelial cell disorder; pain disorder;  
XX KW haematopoietic disorder; blood vessel disorder; metabolic disorder;  
XX KW liver disorder; platelet disorder; glycoprotease.  
XX  
XX  
XX Homo sapiens.  
XX  
XX  
XX US2003009017-A1.  
XX  
XX 09-JAN-2003.  
XX  
XX  
XX 08-NOV-2001; 2001US-00012140.  
XX  
XX  
XX 08-NOV-2000; 2000US-0246768P.  
XX 08-NOV-2000; 2000US-0246772P.  
XX 15-NOV-2000; 2000US-0249185P.  
XX  
XX (LEIBY) LEIBY K R.  
XX (KAP) KAPPELLER-LIBERMANN R.  
XX (GLUC/) GLUCKSMANN M A.  
XX  
XX  
XX Leiby KR, Kapeller-Libermann R, Glucksmann MA;  
XX  
XX WPI; 2003-428888/40.  
XX  
XX N-PSDB; A6A60887.  
XX  
XX New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid  
XX PT molecules, useful for diagnosing, treating cancer, pain, or immune,  
XX PT heart, endothelial cell, hemopoietic, blood vessel, brain, metabolic  
XX PT and liver disorders.  
XX  
XX  
XX Claim 4; Fig 8; 90pp; English.  
XX  
XX  
XX The invention relates to an isolated 38650 (encoding adenosine  
XX CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7  
XX CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or  
XX CC a sequence which is at least 60% identical to the six nucleic acids or  
XX CC their open reading frames, fragments of at least 15 nucleotides,  
XX CC naturally occurring variants, or a DNA insert of the plasmid deposited  
XX CC with the American Type Culture Collection as Accession No. not defined in  
XX CC the specification, which encodes the amino acid sequence). Also included  
XX CC are a host cell containing the nucleic acids (used to produce the

CC proteins), the encoded proteins, an antibody that selectively binds to  
CC the polypeptide, and identifying a compound that binds to/modulates the  
CC activity of the polypeptide. The nucleic acid molecules, polypeptides and  
CC methods are useful for diagnosing, treating cancer, aberrant cellular  
CC proliferation and/or differentiation, immune disorders, heart disorders,  
CC cardiovascular disorders including endothelial cell disorders,  
CC hematopoietic disorders, blood vessel disorders, brain disorders, pain  
CC and metabolic disorders, liver disorders and platelet disorders (many  
CC examples of these disorders are given in the specification). The present  
CC sequence is the Human glycoprotease encoded by cDNA 28472  
XX  
SQ Sequence 414 AA;

Query Match 97.3%; Score 1348; DB 6; Length 414;  
Best Local Similarity 97.4%; Pred. No. 4.3e-143;  
Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MEAHLATRLTNKVEFPPLVLLISGHCCLALVQGVSDFLLGKSLDIAPGMDLDRVAR 60  
DB 148 MEAHLATRLTNKVEFPPLVLLISGHCCLALVQGVSDFLLGKSLDIAPGMDLDRVAR 207  
QY 61 LSLIKHECSTMSGKAI EHLAKQGNRPFDIKPLHAKNCDPFTGLQHTVDKIMMK 120  
DB 208 LSLIKHECSTMSGKAI EHLAKQGNRPFDIKPLHAKNCDPFTGLQHTVDKIMMK 267  
QY 121 EKEEGIEKGQILSSAADIATVQHTMACLVKTRTRAILFCRKORDLLPNNNAVVASGV 180  
DB 268 EKEEGIEKGQILSSAADIATVQHTMACLVKTRTRAILFCRKORDLLPNNNAVVASGV 327  
QY 181 ASNFYIRRALBITLNATQCTLLCPPRLCTDNGIMIANNGIERLRAGIILHIDIGIRYE 240  
DB 328 ASNFYIRRALBITLNATQCTLLCPPRLCTDNGIMIANNGIERLRAGIILHIDIGIRYE 387  
QY 241 PKCPLGVDISKVEGASIKVPOLKMEI 267  
DB 388 PKCPLGVDISKVEGASIKVPOLKMEI 414

## RESULT 8

ADA54471  
ID ADA54471 standard; protein, 364 AA.

AC ADA54471;  
XX  
DT 20-NOV-2003 (first entry)  
DE Human protein, SEQ ID 2039.  
XX  
KM Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Noctropic;  
KM Gene Therapy; human; secretory protein; membrane proteins; cancer;  
KM inflammatory disease; osteoporosis; neurological disease.  
XX  
OS Homo sapiens.  
XX  
PN EP1293569-A2.  
XX  
PD 19-MAR-2003.  
XX  
PF 21-MAR-2002; 2002EP-00006586.  
XX  
PR 14-SEP-2001; 2001JP-00328381.  
PR 24-JAN-2002; 2002US-0350435P.  
PA (HELI-) HELIX RES INST.  
PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
PI Iisogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yamamoto U, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
PI Seti N, Yoshikawa T, Otsuka M, Nagahari K, Maehno Y;  
XX WPI; 2003-395539/38.  
XX DR N-PSDB; ADA52832.  
XX

PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
PT and/or membrane proteins, useful for developing medicines for diseases in  
PT which the gene is involved, or as target molecules for gene therapy.  
XX  
PS Claim 14; SEQ ID NO 2039; 205pp; English.

XX The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
SQ Sequence 364 AA;

Query Match 79.8%; Score 1105; DB 6; Length 364;  
Best Local Similarity 98.6%; Pred. No. 1.1e-115;  
Matches 214; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MEAHLATRLTNKVEFPPLVLLISGHCCLALVQGVSDFLLGKSLDIAPGMDLDRVAR 60  
DB 148 MEAHLATRLTNKVEFPPLVLLISGHCCLALVQGVSDFLLGKSLDIAPGMDLDRVAR 207  
QY 61 LSLIKHECSTMSGKAI EHLAKQGNRPFDIKPLHAKNCDPFTGLQHTVDKIMMK 120  
DB 208 LSLIKHECSTMSGKAI EHLAKQGNRPFDIKPLHAKNCDPFTGLQHTVDKIMMK 267  
QY 121 EKEEGIEKGQILSSAADIATVQHTMACLVKTRTRAILFCRKORDLLPNNNAVVASGV 180  
DB 268 EKEEGIEKGQILSSAADIATVQHTMACLVKTRTRAILFCRKORDLLPNNNAVVASGV 327  
QY 181 ASNFYIRRALBITLNATQCTLLCPPRLCTDNGIMIA 217  
DB 328 ASNFYIRRALBITLNATQCTLLCPPRLCTDNGIMIA 364

## RESULT 9

ABB69133  
ID ABB69133 standard; protein, 409 AA.

AC ABB69133;  
XX  
DT 26-MAR-2002 (first entry)  
DE Drosophila melanogaster polypeptide SEQ ID NO 34191.  
XX  
KM Drosophila; developmental biology; cell signalling; insecticide;  
KM pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX DR N-PSDB; ABL13236.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 34191, 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
CC sequences (AB101840-AB116175) and the encoded proteins (AB57737-  
CC AB572072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 409 AA;  
  
Query Match 30.0%; Score 415.5; DB 4; Length 409;  
Best Local Similarity 38.2%; Pred. No. 1.4e-37;  
Matches 100; Conservative 43; Mismatches 104; Indels 15; Gaps 6;  
  
QY 1 MEAHALTRLTN--KVEPPVLVLTSGGHCCLAVGVSDPFLKSLDIAPGMDLKVKA 58  
DB 136 MEAHALQARMHPEQIGYFPLCLASGGHCQLVANGRRLTLTQTLDDAEGEAPDKIG 195  
QY 59 RLSLIKHPECSTMSGKAIENHLAK--QGNRFHFDIKRPLHAKNCDPFTGLQHTDKII 117  
DB 196 RLRLHLPLRYRLNMNGRAIENHQAASDPLAVERPPLAQQRNCFSPAGIKNNSPRAI 255  
QY 118 MKKEKEGIEKGQILSSADIAATVQHTMACHLVKTRRAILFC--KQRIILPQNNAVLV 175  
DB 256 RARERARTPPDGVISNYGDECAGILRSVSRHLMRTORALEYCLLPHQLEPDPPTLV 315  
QY 176 ASGGVAAENFYIRBALLETITNAQCTLLCPPRLCTDNGIMIAVNGIERLRAGLILHDE 235  
DB 316 MSGGVANNDAIYANIEHLAAQYGCSPFRSKRYCSNDGVMIAHMGVEQL-----LDQKE 369  
QY 236 -GIRPEKCPGLVDISKEVGEA 256  
DB 370 ASTRYDYD---SIDIQSAGFA 388  
  
RESULT 10  
AAV52216  
ID AAV52216 standard; protein; 463 AA.  
XX  
AC AAV52216;  
XX  
DT 09-FEB-2000 (first entry)  
XX  
DE Arabidopsis thaliana yjD protein homologue.  
XX  
KW yjD protein; essential; Gram positive; Gram negative; conserved; motif;  
KW identification; antagonist; antibacterial; antibiotic; broad spectrum;  
KW treatment; infection; resistance; drug target.  
XX  
OS Arabidopsis thaliana.  
XX  
FH Key Location/Qualifiers  
FT Region 86..96  
FT Region /note= "yjD conserved motif 3"  
FT Region 111..131  
FT Region /note= "yjD conserved motif 4"  
FT Region 152..198  
FT Region /note= "yjD conserved motif 2"  
FT Region 208..259  
FT /note= "yjD conserved motif 1"  
XX  
PN MO954470-A2.  
XX  
PD 28-OCT-1999.  
XX  
PP 20-APR-1999; 99WO-EP002635.  
XX  
PN 22-APR-1999; 36GB-00006423.  
XX  
PA (GLAX ) GLAXO GROUP LTD.  
XX

PI Arigoni F, Edgeron MD, Lofèrer H, Peitsch MC;  
XX  
DR WPI; 2000-013253/01.  
XX  
PT Novel bacterial polypeptides used to identify broad spectrum antibiotics.  
XX  
PS Claim 1, Fig 1, 55pp; English.  
XX  
CC Proteins AAV52202-Y52217 encompass a novel family of proteins designated  
CC the yjD family, after the name given to the *Escherichia coli* family  
CC member. These proteins are essential for the survival of both Gram  
CC negative and Gram positive bacteria, although no function has as yet been  
CC ascribed to these proteins. The yjD proteins, fragments of yjD motifs  
CC (for example, fragments encompassing one or more conserved yjD motifs  
CC such as AAV52218-Y52284) and nucleotides encoding them can be used to  
CC identify antagonists and broad spectrum antibacterial compounds. These  
CC antagonists and compounds can be used to treat a wide range of bacterial  
CC infections. New antibiotics are urgently needed, as serious bacterial  
CC infections and antibiotic resistant strains are becoming increasingly  
CC prevalent. The proteins of the invention are essential proteins for  
CC bacterial viability, and represent new targets for antibiotics  
XX  
SQ Sequence 463 AA;  
  
Query Match 29.6%; Score 409.5; DB 3; Length 463;  
Best Local Similarity 36.6%; Pred. No. 8.2e-37;  
Matches 100; Conservative 37; Mismatches 91; Indels 45; Gaps 6;  
  
QY 1 MEAHALTRLTN--KVEPPVLVLTSGGHCCLAVGVSDPFLKSLDIAPGMDLKVKA 59  
DB 192 MEAHALVALVBOELSPFMALLISGHNILVLAHLCAGYQYGLTVDDAIGEAFFDKAK 251  
QY 60 RLSLIKHPECSTMSGKAIENHLAKQGNRFHFDIKRPLHAKNCDPFTGLQHTDKIIM 119  
DB 252 WGLDMH-----RSGPFAVEELALBDAKSVKFNVMKYKNDPFSYAGLKTQVRLAIEA 306  
QY 120 KEKEGIEKGQILSSADIAATVQHTMACHLVKTRRAILFCQRIILPQNNAVLVASG 179  
DB 307 KE-----IRRADIASFQRAVAVLHEKCRALDMLA---LEPSIGHVYISGG 353  
QY 180 VASNFYIRBALLETITNAQCTLLCPPRLCTDNGIMIAVNGIERLRAGLILHDEGIN 239  
DB 354 VASNKYVRLNNIYENKMLKIVCPPSLCTDNGVAVMTGLSEHFVVG-----RY 403  
QY 240 E-----PKCPGLVDISKEVGEA 256  
DB 404 DPPPATEPEDEVYDLRPRWPLGEBEYAKGRSEA 436  
  
RESULT 11  
AAG19287  
ID AAG19287 standard; protein; 245 AA.  
XX  
AC AAG19287;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 21031.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PP 25-FEB-2000; 2000EP-00301439.  
XX  
PN 25-FEB-1999; 99US-0121825P.  
XX  
PR 05-MAR-1999; 99US-012180P.  
XX

PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 18-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137750P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138549P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140822P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 22-JUL-1999; 99US-0145193P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-014819P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148655P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 20-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 23-AUG-1999; 99US-0150566P.  
PR 25-AUG-1999; 99US-0150884P.  
PR 26-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-015659P.  
PR 28-SEP-1999; 99US-015648P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.

PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158023P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159328P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 16-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 26-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161931P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 28.8%; Score 398.5; DB 3; Length 245;  
Best Local Similarity 38.7%; Pred. No. 5; 6e-36;  
Matches 98; Conservative 34; Mismatches 90; Indels 31; Gaps 6;

QY 1 MEAHATITRLT-NKVEPFVLVLLISGCHLALVGVSDPFLILGKSLDIAPGMDLKVVR 59  
D 1 MEATIVARLVEGELSPFFMALISGCHNLVLAKHQLQYQLGTTVDAIGEARDKTKAK 60  
QY 60 RLSIIRKPECSMTSGGKAIEHLAKQGNRFHDIKPLHAKNKNDSFPGLOHVTDKIIMK 119  
D 61 WIGEDMH-----RSGPRAVEELALEGDAKSVKFNVPKMKYKHCNFSYAGLKTQVLAIEA 115  
QY 120 KEKEGIEKQIILSA-----ADIAATQHTMACILVKTTHAILFCQRDLIPQNN 171  
D 116 KE-----IDAKCPVSSATNEDRRNRADIAASFQVAVLHLEKCERAIDMALE---LEPSI 168  
QY 172 AVLVASGVASNFYIRRLAILITNAIOCTLLCPRLCTDNGIMIAMNGIERLRAGLGL 231  
D 169 KHWIISGVASNKYVRLRLNINVENKNIKLVCPPLSLCTDNGVAMVGLHFRVVG---- 224  
QY 232 HDIEGIRYEPKCP 244  
D 225 -----RYDPEPP 231

RESULT 12  
AAG19286  
ID AAG19286 standard; protein; 439 AA.

XX AAG19286;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 21030.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW Hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.  
XX PD 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-00301439.  
XX 25-FEB-1999; 99US-0121825P.  
XX 05-MAR-1999; 99US-0123180P.  
XX 09-MAR-1999; 99US-0123548P.  
XX 23-MAR-1999; 99US-0125788P.  
XX 25-MAR-1999; 99US-0126264P.  
XX 29-MAR-1999; 99US-0126785P.  
XX 01-APR-1999; 99US-0127462P.  
XX 06-APR-1999; 99US-0128234P.  
XX 08-APR-1999; 99US-0128714P.  
XX 16-APR-1999; 99US-0129845P.  
XX 19-APR-1999; 99US-0130077P.  
XX 21-APR-1999; 99US-0130449P.  
XX 23-APR-1999; 99US-0130510P.  
XX 28-APR-1999; 99US-0130891P.  
XX 30-APR-1999; 99US-0131449P.  
XX 30-APR-1999; 99US-0132048P.  
XX 04-MAY-1999; 99US-0132407P.  
XX 05-MAY-1999; 99US-0132484P.  
XX 06-MAY-1999; 99US-0132485P.  
XX 06-MAY-1999; 99US-0132486P.  
XX 07-MAY-1999; 99US-0132487P.  
XX 11-MAY-1999; 99US-0132863P.  
XX 14-MAY-1999; 99US-0134256P.  
XX 14-MAY-1999; 99US-0134218P.  
XX 14-MAY-1999; 99US-0134219P.  
XX 14-MAY-1999; 99US-0134370P.  
XX 18-MAY-1999; 99US-0134768P.  
XX 19-MAY-1999; 99US-0134941P.  
XX 20-MAY-1999; 99US-0135124P.  
XX 21-MAY-1999; 99US-0135333P.  
XX 24-MAY-1999; 99US-0135629P.  
XX 25-MAY-1999; 99US-0136021P.  
XX 27-MAY-1999; 99US-0136392P.  
XX 28-MAY-1999; 99US-0136782P.  
XX 01-JUN-1999; 99US-0137222P.  
XX 03-JUN-1999; 99US-0137528P.  
XX 04-JUN-1999; 99US-0137502P.  
XX 07-JUN-1999; 99US-0137724P.  
XX 08-JUN-1999; 99US-0138094P.  
XX 10-JUN-1999; 99US-0138540P.  
XX 10-JUN-1999; 99US-0138847P.  
XX 14-JUN-1999; 99US-0139119P.  
XX 16-JUN-1999; 99US-0139452P.  
XX 17-JUN-1999; 99US-0139453P.  
XX 17-JUN-1999; 99US-0139492P.  
XX 18-JUN-1999; 99US-0139455P.  
XX 18-JUN-1999; 99US-0139456P.  
XX 18-JUN-1999; 99US-0139457P.  
XX 18-JUN-1999; 99US-0139458P.  
XX 18-JUN-1999; 99US-0139459P.  
XX 18-JUN-1999; 99US-0139460P.  
XX 18-JUN-1999; 99US-0139461P.  
XX 18-JUN-1999; 99US-0139462P.  
XX 18-JUN-1999; 99US-0139463P.  
XX 18-JUN-1999; 99US-0139750P.  
XX 18-JUN-1999; 99US-0139763P.  
XX 21-JUN-1999; 99US-0139817P.  
XX 22-JUN-1999; 99US-0139899P.  
XX 23-JUN-1999; 99US-0140353P.  
XX 23-JUN-1999; 99US-0140354P.  
XX 24-JUN-1999; 99US-0140695P.  
XX 28-JUN-1999; 99US-0140823P.  
XX 29-JUN-1999; 99US-0140991P.  
XX 30-JUN-1999; 99US-0141287P.



PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144864P.  
PR 20-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0148171P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149425P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.

PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158331P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0161040P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 28.8%; Score 398.5; DB 3; length 439;  
Best Local Similarity 38.7%; Pred. No. 1.3e-35;  
Matches 98; Conservative 34; Mismatches 90; Indels 31; Gaps 6;

QY 1 MEAHALTRLT-NKVEPPLVILISGRCILAVOGVSPFLIGKSLDIAPEMDLKVAR 59  
DB 195 MEAHTLVARLVEQELSPFMALLISGHNILVLAHKLCQYTOIGTVDDAIGRAPDKTAK 254  
QY 60 RLSLIKHPKPCSTMSGKALIEHLAKOGNRFHFDIKPRLHHAKNDPESFTGLOHTDKITMK 119  
DB 255 WGLDMH-----RSGGPAVEELLEGDAKSVKFNVMKHKNDKFNFGAGLKTQVRLAIEA 309  
QY 120 KEKEBIEKQILISSA-----ADIATVOHTMACHVKTHTRAILPFCKGRDLIPONN 171  
DB 310 KE-----IDAKCPVSSATNEDRRKADIIAASFQVAVALHBEKERRIDALE---LEPSI 362  
QY 172 AVLVASGVASNFYIRALEILTNAQTLLCPPLCTDNGIMIAMNGIERLPAIGLIL 231  
DB 363 KHMVISGVASNKYVALRLNNIIVENKULKIVCPPSLCTDNGVMVAMTGLEHFRVG---- 418  
QY 232 HDIEGIRYEPKCP 244  
DB 419 -----RYDPPPP 425

RESULT 13  
AAG19285  
ID AAG19285 standard; protein; 444 AA.  
XX  
AC AAG19285;  
XX

DT 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 21029.  
DE  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX Arabidopsis thaliana.  
OS  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 06-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137529P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143524P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 18-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.

PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157533P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	29-OCT-1999;	99US-0162142P.

```

Query Match 28.8% Score 398.5 DB 3 Length 444;
Best Local Similarity 38.7% Pred No. 14e-35;
Matches 98; Conservative 34; Mismatches 90; Indels 31; Gaps 6

OY 1 MEAHALITRLT-NKVEPPELVLLISGGHCLALVQSVDSFDLLIGSLDIAPGDMIDKVAR 59
    |||||:::||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 200 MEAHITVARIVQELSPFPFMAIIISGHNILVLANKIGQYQLGTGVDDAIGBAFDKPAK 259
    |||||:::||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 60 RLSLIKPEPCSTMGGCKAIENHLAKGNRPHDIKRPYLHAKKCNDSFTGLQHYTDKILMK 119
    |||||:::||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 260 WLGLDMH-----RSGGPAVEELALGDAKSVFENYPMKXKHCNPFSAIGLQVRLAIEA 314
    |||||:::||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 120 KEKEKIEKGQILLSA-----ADIAATVQHTMACHLYKRTTHRIILCKQKRDLLPQNN 171
    |||||:::||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 315 KE---IDAKCPVSSATVEDRRNRNDIASFORVAVLHEEKCEKRIADALE---LEPSI 367
    |||||:::||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 172 AVLVASGVASNFYIRRALEILITNATQCTLLCPPEPLCTDNGIMIMANGIEERLAGLIL 231
    |||||:::||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 368 KHMVTSGGVASKRYRLRLNMIIVENKNLKVCCPPELCTDNGVMAMTGLEHRRVG---- 423
    |||||:::||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 232 HDIEGIRYEPKCP 244
    |||||:::||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

[illegible]



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 16:43:58 ; Search time 70.5228 Seconds

(without alignments)  
1584.102 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414

Perfect score: 1385

Sequence: 1 MEAHALITRLTNKVEPPPLV.....DISKEVGASIKVQLKMEI 267

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867879 seqs, 418409474 residues

Total number of hits satisfying chosen parameters: 1867879

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep.\*

2: /cgn2\_6/ptodata/1/pubppaa/PCT\_NEW\_PUB.pep.\*

3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep.\*

4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep.\*

5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep.\*

6: /cgn2\_6/ptodata/1/pubppaa/PCTUS\_PUBCOMB.pep.\*

7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep.\*

8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep.\*

9: /cgn2\_6/ptodata/1/pubppaa/US09A\_PUBCOMB.pep.\*

10: /cgn2\_6/ptodata/1/pubppaa/US09B\_PUBCOMB.pep.\*

11: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pep.\*

12: /cgn2\_6/ptodata/1/pubppaa/US10A\_NEW\_PUB.pep.\*

13: /cgn2\_6/ptodata/1/pubppaa/US10A\_PUBCOMB.pep.\*

14: /cgn2\_6/ptodata/1/pubppaa/US10B\_PUBCOMB.pep.\*

15: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep.\*

16: /cgn2\_6/ptodata/1/pubppaa/US10D\_PUBCOMB.pep.\*

17: /cgn2\_6/ptodata/1/pubppaa/US10E\_PUBCOMB.pep.\*

18: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep.\*

19: /cgn2\_6/ptodata/1/pubppaa/US11A\_PUBCOMB.pep.\*

20: /cgn2\_6/ptodata/1/pubppaa/US11\_NEW\_PUB.pep.\*

21: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep.\*

22: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1385	100.0	267	US-10-067-443-22	Sequence 22, Appl
2	1385	100.0	267	US-10-649-273-22	Sequence 22, Appl
3	1385	100.0	267	US-10-651-722-22	Sequence 22, Appl
4	1385	100.0	414	US-10-067-443-2	Sequence 2, Appl
5	1385	100.0	414	US-10-649-273-2	Sequence 2, Appl
6	1385	100.0	414	US-10-651-722-2	Sequence 2, Appl
7	1385	100.0	414	US-10-480-988-8	Sequence 8, Appl
8	1358.5	98.1	439	US-10-649-273-19	Sequence 19, Appl
9	1358.5	98.1	439	US-10-651-722-19	Sequence 19, Appl
10	1348	97.3	414	US-10-012-140-5	Sequence 5, Appl

12	1105	79.8	364	15	US-10-094-749-2039	Sequence 2039, Ap
13	422.5	30.5	445	15	US-10-424-599-209259	Sequence 209259,
14	415.5	30.0	409	20	US-11-097-143-34191	Sequence 34191, A
15	409.5	29.6	463	14	US-10-067-443-3	Sequence 3, Appl
16	409.5	29.6	463	15	US-10-649-273-3	Sequence 3, Appl
17	409.5	29.6	463	15	US-10-651-722-3	Sequence 3, Appl
18	359	25.9	444	16	US-10-437-965-113732	Sequence 113732,
19	342	24.7	251	14	US-10-081-051-9	Sequence 9, Appl
20	321.5	23.2	382	15	US-10-282-122A-50858	Sequence 50858, A
21	317	22.9	333	14	US-10-012-140-25	Sequence 25, Appl
22	314.5	22.7	312	15	US-10-282-122A-44499	Sequence 44499, A
23	308.5	22.3	348	15	US-10-282-122A-63156	Sequence 63156, A
24	304.5	22.0	343	15	US-10-282-122A-67227	Sequence 67227, A
25	294	21.2	347	14	US-10-012-140-24	Sequence 24, Appl
26	291.5	21.0	341	9	US-09-815-242-11798	Sequence 11798, A
27	291.5	21.0	341	15	US-10-282-122A-66200	Sequence 66200, A
28	291.5	21.0	341	18	US-10-958-216-50	Sequence 52, Appl
29	289.5	20.9	341	18	US-10-958-216-52	Sequence 11043, A
30	289.5	20.9	342	9	US-09-815-242-11043	Sequence 58204, A
31	289.5	20.9	342	15	US-10-282-122A-58204	Sequence 68438, A
32	280.5	20.3	340	15	US-10-282-122A-68438	Sequence 13780, A
33	279.5	20.2	337	9	US-09-815-242-13780	Sequence 75485, A
34	279.5	20.2	337	15	US-10-282-122A-75485	Sequence 78161, A
35	279.5	20.2	337	15	US-10-282-122A-78161	Sequence 55404, A
36	277.5	20.0	335	15	US-10-282-122A-55404	Sequence 61938, A
37	277	20.0	309	15	US-10-282-122A-61938	Sequence 67933, A
38	277	20.0	341	15	US-10-282-122A-67933	Sequence 4, Appl
39	276	19.9	421	14	US-10-067-443-4	Sequence 28, Appl
40	276	19.9	421	15	US-10-649-273-4	Sequence 4, Appl
41	276	19.9	421	15	US-10-649-273-28	Sequence 28, Appl
42	276	19.9	421	15	US-10-651-722-4	Sequence 4, Appl
43	276	19.9	421	15	US-10-651-722-28	Sequence 28, Appl
44	276	19.9	421	15	US-10-651-722-28	Sequence 28, Appl
45	275.5	19.9	337	9	US-09-815-242-10304	Sequence 10304, A

## ALIGNMENTS

RESULT 1

US-10-067-443-22

Sequence 22, Application US/10067443

Publication No. US20030082782A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN

FILE REFERENCE: D0073 NP

CURRENT APPLICATION NUMBER: US/10/067,443

CURRENT FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.0

SEQ ID NO 22

LENGTH: 267

TYPE: PRT

ORGANISM: homo sapiens

US-10-067-443-22

Query Match 100.0%; Score 1385; DB 14; Length 267;

Best Local Similarity 100.0%; Pred. No. 2.2e-139; Mismatches 0; Gaps 0;

Matches 267; Conservative 0; Indels 0;

QY 1 MEAHALITRLTNKVEPPPLVLLISGGHCLALVOGSDPLLGKSLDIARGMDLVARR 60

DB 1 MEAHALITRLTNKVEPPPLVLLISGGHCLALVOGSDPLLGKSLDIARGMDLVARR 60

QY 61 LSLRHPESCTSGSKAIEHLAKQGNRPFDIKPPLAHKNCDFSTGQHYTDKIMK 120

DB 61 LSLRHPESCTSGSKAIEHLAKQGNRPFDIKPPLAHKNCDFSTGQHYTDKIMK 120



```
QY      121 EKEEGIEKGQILSSAADIAATVQHTMACHLVKTRTHRALIFCKORDLLPQNNAVLVASGCV 180
      |||
      268 EKEEGIEKGQILSSAADIAATVQHTMACHLVKTRTHRALIFCKORDLLPQNNAVLVASGCV 327
Db
QY      181 ASNFYIRRALEILTNAQTCTLLCPPRLCTDNGIMIANWGIERLRAGIGIHDIEGIRYE 240
      |||
      328 ASNFYIRRALEILTNAQTCTLLCPPRLCTDNGIMIANWGIERLRAGIGIHDIEGIRYE 387
Db
QY      241 PKCPLGVDISKVEGASIKVPOLKMEI 267
      |||
      388 PKCPLGVDISKVEGASIKVPOLKMEI 414
Db
```

## RESULT 5

```
US-10-649-273-2
; Sequence 2, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649, 273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266, 518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067, 443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282, 814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-649-273-2
```

```
Query Match      100.0%; Score 1385; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 4.2e-139;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 MEAHALTIRLTNKVEPFLVLLISGHCILALVQGVSDPFLIGKSLDIAPGMDLVARR 60
      |||
      148 MEAHALTIRLTNKVEPFLVLLISGHCILALVQGVSDPFLIGKSLDIAPGMDLVARR 207
Db
QY      61 LSLIKPECSITSGGKAI EHLAKQGNRFHDIKPLHAKNCDPSTGLOHTVDKTIIMKK 120
      |||
      208 LSLIKPECSITSGGKAI EHLAKQGNRFHDIKPLHAKNCDPSTGLOHTVDKTIIMKK 267
Db
QY      121 EKEEGIEKGQILSSAADIAATVQHTMACHLVKTRTHRALIFCKORDLLPQNNAVLVASGCV 180
      |||
      268 EKEEGIEKGQILSSAADIAATVQHTMACHLVKTRTHRALIFCKORDLLPQNNAVLVASGCV 327
Db
QY      181 ASNFYIRRALEILTNAQTCTLLCPPRLCTDNGIMIANWGIERLRAGIGIHDIEGIRYE 240
      |||
      328 ASNFYIRRALEILTNAQTCTLLCPPRLCTDNGIMIANWGIERLRAGIGIHDIEGIRYE 387
Db
QY      241 PKCPLGVDISKVEGASIKVPOLKMEI 267
      |||
      388 PKCPLGVDISKVEGASIKVPOLKMEI 414
Db
```

## RESULT 6

```
US-10-651-722-2
; Sequence 2, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651, 722
; PRIOR FILING DATE: 2003-08-29
```

```
; PRIOR APPLICATION NUMBER: US 60/266, 518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067, 443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282, 814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-651-722-2
```

```
Query Match      100.0%; Score 1385; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 4.2e-139;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 MEAHALTIRLTNKVEPFLVLLISGHCILALVQGVSDPFLIGKSLDIAPGMDLVARR 60
      |||
      148 MEAHALTIRLTNKVEPFLVLLISGHCILALVQGVSDPFLIGKSLDIAPGMDLVARR 207
Db
QY      61 LSLIKPECSITSGGKAI EHLAKQGNRFHDIKPLHAKNCDPSTGLOHTVDKTIIMKK 120
      |||
      208 LSLIKPECSITSGGKAI EHLAKQGNRFHDIKPLHAKNCDPSTGLOHTVDKTIIMKK 267
Db
QY      121 EKEEGIEKGQILSSAADIAATVQHTMACHLVKTRTHRALIFCKORDLLPQNNAVLVASGCV 180
      |||
      268 EKEEGIEKGQILSSAADIAATVQHTMACHLVKTRTHRALIFCKORDLLPQNNAVLVASGCV 327
Db
QY      181 ASNFYIRRALEILTNAQTCTLLCPPRLCTDNGIMIANWGIERLRAGIGIHDIEGIRYE 240
      |||
      328 ASNFYIRRALEILTNAQTCTLLCPPRLCTDNGIMIANWGIERLRAGIGIHDIEGIRYE 387
Db
QY      241 PKCPLGVDISKVEGASIKVPOLKMEI 267
      |||
      388 PKCPLGVDISKVEGASIKVPOLKMEI 414
Db
```

## RESULT 7

```
US-10-480-988-8
; Sequence 8, Application US/10480988
; Publication No. US20050069877A1
; GENERAL INFORMATION:
; APPLICANT: GANDHI, Ameena R.; KABLE, Amy B.;
; APPLICANT: SMARNAKAR, Anita; HAPALA, Ariti J.A.;
; APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;
; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;
; APPLICANT: HONCHELL, Cynthia D.; NGUYEN, Daniel B.;
; APPLICANT: LU, Dzung Aina M.; LEE, Ernestine A.;
; APPLICANT: YUE, Henry; FORSYTHE, Ian J.;
; APPLICANT: BARROSO, Ines; RAMKUMAR, Jayalakshmi;
; APPLICANT: GRIFFIN, Jennifer A.; LI, Joana X.;
; APPLICANT: YANG, Junming; THANGAVELU, Kavitha;
; APPLICANT: GIERZEN, Kimberly J.; DING, Li;
; APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.;
; APPLICANT: YAO, Monique G.; CHAKLA, Natinder K.;
; APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
; APPLICANT: LEB, Sally; BECHA, Shanya D.;
; APPLICANT: LEB, Soo Yeun; TRAN, Uyen K.;
; APPLICANT: ELIOTT, Vicki S.; LUO, Wen;
; APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;
; APPLICANT: LU, Yan; ZBARAJADIAN, Yeganeh
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PF-1040 USN
; CURRENT APPLICATION NUMBER: US/10/480, 988
; PRIOR FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: PCT/US02/19360
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/300, 508
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/303, 445
; PRIOR FILING DATE: 2001-07-06
```



PRIOR APPLICATION NUMBER: US 60/305,405  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/311,442  
PRIOR FILING DATE: 2001-08-09  
PRIOR APPLICATION NUMBER: US 60/314,821  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/315,992  
PRIOR FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: US 60/378,205  
PRIOR FILING DATE: 2002-05-03  
NUMBER OF SEQ ID NOS: 56  
SOFTWARE: PERL Program  
SEQ ID NO 8  
LENGTH: 414  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE: misc\_feature  
NAME/KEY: misc\_feature  
OTHER INFORMATION: Incyte ID No: 7632424CD1  
US-10-480-988-8

Query Match 100.0%; Score 1385; DB 17; Length 414;  
Best Local Similarity 100.0%; Pred. No. 4.2e-139;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEAHALTIRLTNKVEPPLVLLISGGHCLALVOGVSDPFLLLGKSLDIAFGMDLVARR 60  
Db 148 MEAHALTIRLTNKVEPPLVLLISGGHCLALVOGVSDPFLLLGKSLDIAFGMDLVARR 207  
Qy 61 LSLIKPECSSTMSGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLOHTVDKTIIMKK 120  
Db 208 LSLIKPECSSTMSGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLOHTVDKTIIMKK 267  
Qy 121 EKEEGIKGQILSSAADIAATVOHTMACHLVKTRTALIFCKQRDLPPONNAVIVASGV 180  
Db 268 EKEEGIKGQILSSAADIAATVOHTMACHLVKTRTALIFCKQRDLPPONNAVIVASGV 327  
Qy 181 ASNFYIRALEITLNATQCTLLCPPLCTDNGMIANNGIERLHAGILHDIIGIRYE 240  
Db 328 ASNFYIRALEITLNATQCTLLCPPLCTDNGMIANNGIERLHAGILHDIIGIRYE 387  
Qy 241 PKCPLGVDSKEVGEASIKVQLKMEI 267  
Db 388 PKCPLGVDSKEVGEASIKVQLKMEI 414

RESULT 8  
US-10-067-443-19  
; Sequence 19, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 19  
; LENGTH: 439  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-067-443-19

Query Match 98.1%; Score 1358.5; DB 14; Length 439;  
Best Local Similarity 91.1%; Pred. No. 3.2e-136;  
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

Qy 1 MEAHALTIRLTNKVEPPLVLLISGGHCLALVOGVSDPFLLLGKSLDIAFGMDLVARR 60  
Db 148 MEAHALTIRLTNKVEPPLVLLISGGHCLALVOGVSDPFLLLGKSLDIAFGMDLVARR 207  
Qy 61 LSLIKPECSSTMSGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLOHTVDKTIIMKK 120  
Db 208 LSLIKPECSSTMSGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLOHTVDKTIIMKK 267  
Qy 121 EKEEGIKGQILSSAADIAATVOHTMACHLVKTRTALIFCKQRDLPPONNAVIVASGV 180  
Db 268 EKEEGIKGQILSSAADIAATVOHTMACHLVKTRTALIFCKQRDLPPONNAVIVASGV 327  
Qy 181 ASNFYIRALEITLNATQCTLLCPPLCTDNGMIANNGIERLHAGILHDIIGIRYE 240  
Db 328 ASNFYIRALEITLNATQCTLLCPPLCTDNGMIANNGIERLHAGILHDIIGIRYE 387  
Qy 241 PKCPLGVDSKEVGEASIKVQLKMEI 267  
Db 388 PKCPLGVDSKEVGEASIKVQLKMEI 439

RESULT 9  
US-10-649-273-19  
; Sequence 19, Application US/10649273  
; Publication No. US20040043407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649,273  
; CURRENT FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 19  
; LENGTH: 439  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-649-273-19

Query Match 98.1%; Score 1358.5; DB 15; Length 439;  
Best Local Similarity 91.1%; Pred. No. 3.2e-136;  
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

Qy 1 MEAHALTIRLTNKVEPPLVLLISGGHCLALVOGVSDPFLLLGKSLDIAFGMDLVARR 60  
Db 148 MEAHALTIRLTNKVEPPLVLLISGGHCLALVOGVSDPFLLLGKSLDIAFGMDLVARR 207  
Qy 61 LSLIKPECSSTMSGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLOHTVDKTIIMKK 120  
Db 208 LSLIKPECSSTMSGKAI EHLAKQGNRFHDIKPLHAKNCDPFTGLOHTVDKTIIMKK 267  
Qy 121 EKEEGIKGQILSSAADIAATVOHTMACHLVKTRTALIFCKQRDLPPONNAVIVASGV 180  
Db 268 EKEEGIKGQILSSAADIAATVOHTMACHLVKTRTALIFCKQRDLPPONNAVIVASGV 327  
Qy 181 ASNFYIRALEITLNATQCTLLCPPLCTDNGMIANNGIERLHAGILHDIIGIRYE 240  
Db 328 ASNFYIRALEITLNATQCTLLCPPLCTDNGMIANNGIERLHAGILHDIIGIRYE 387  
Qy 241 PKCPLGVDSKEVGEASIKVQLKMEI 267  
Db 388 PKCPLGVDSKEVGEASIKVQLKMEI 439

RESULT 10  
US-10-651-722-19  
; Sequence 19, Application US/10651722

```
Publication No. US20040048302A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
FILE REFERENCE: D0073 DIV
CURRENT APPLICATION NUMBER: US/10/651,722
PRIOR FILING DATE: 2003-08-29
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 10/067,443
PRIOR FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.2
SEQ ID NO 19
LENGTH: 439
TYPE: PRT
ORGANISM: homo sapiens
US-10-651-722-19

Query Match          98.1%; Score 1358.5; DB 15; Length 439;
Best Local Similarity 91.1%; Pred. No. 3.2e-136;
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

Qy 1 MEAHLTIRLTNKKVEFPFLVLLISGHCCLALVOGVSDFLLGKSLDIAPGMDLKVARR 60
Db 148 MEAHLTIRLTNKKVEFPFLVLLISGHCCLALVOGVSDFLLGKSLDIAPGMDLKVARR 207
Qy 61 LSLIKHPECSTMSGKAIIEHLAKQGNRFHFDIKPPLHAKNCDFSFTGLQHTVDKIMKK 120
Db 208 LSLIKHPECSTMSGKAIIEHLAKQGNRFHFDIKPPLHAKNCDFSFTGLQHTVDKIMKK 267
Qy 121 EKEEGI-----EKGQIISADIAATVQHTMACHLVYKRTK 155
Db 268 EKEEGIFLISKVEQINPGLCLKIAHPCRYEKQIISADIAATVQHTMACHLVYKRTK 327
Qy 156 RALIFCQRDLTPONNAVTVASGVASNFYIRALLETITNAOCTLCPPRLCTDNGIM 215
Db 328 RALIFCQRDLTPONNAVTVASGVASNFYIRALLETITNAOCTLCPPRLCTDNGIM 387
Qy 216 IAWNGIERLAGLIDIEGIRYEPKPLGVDISKVEGASIKVPOLKMEI 267
Db 388 IAWNGIERLAGLIDIEGIRYEPKPLGVDISKVEGASIKVPOLKMEI 439

RESULT 11
US-10-012-140-5
Sequence 5, Application US/10012140
Publication No. US20030009017A1
GENERAL INFORMATION:
APPLICANT: Leibny, Kevin R.
APPLICANT: Kapeller-Libermann, Rosana
APPLICANT: Glucksmann, Maria A.
TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: 381552004900
CURRENT APPLICATION NUMBER: US/10/012,140
PRIOR FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: 60/246,768
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/246,772
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/249,185
PRIOR FILING DATE: 2000-11-15
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 414
TYPE: PRT
ORGANISM: Homo sapiens
US-10-012-140-5
```

```
Query Match          97.3%; Score 1348; DB 14; Length 414;
Best Local Similarity 97.4%; Pred. No. 3.9e-135;
Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MEAHLTIRLTNKKVEFPFLVLLISGHCCLALVOGVSDFLLGKSLDIAPGMDLKVARR 60
Db 148 MEAHLTIRLTNKKVEFPFLVLLISGHCCLALVOGVSDFLLGKSLDIAPGMDLKVARR 207
Qy 61 LSLIKHPECSTMSGKAIIEHLAKQGNRFHFDIKPPLHAKNCDFSFTGLQHTVDKIMKK 120
Db 208 LSLIKHPECSTMSGKAIIEHLAKQGNRFHFDIKPPLHAKNCDFSFTGLQHTVDKIMKK 267
Qy 121 EKEEGIENQIISADIAATVQHTMACHLVYKRTKRALIFCQRDLTPONNAVTVASGV 180
Db 268 KOEGIEKGQIISADIAATVQHTMACHLVYKRTKRALIFCQRDLTPONNAVTVASGV 327
Qy 181 ASNFYIRALLETITNAOCTLCPPRLCTDNGIMIAWNGIERLAGLIDIEGIRY 240
Db 328 ASNFYIRALLETITNAOCTLCPPRLCTDNGIMIAWNGIERLAGLIDIEGIRY 387
Qy 241 PKCPLGVDISKVEGASIKVPOLKMEI 267
Db 388 PKCPLGVDISKVEGASIKVPOLKMEI 414
```

```
RESULT 12
US-10-094-749-2039
Sequence 2039, Application US/10094749
Publication No. US20030219741A1
GENERAL INFORMATION:
APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORI
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHITO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749
PRIOR FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2039
LENGTH: 364
TYPE: PRT
ORGANISM: Homo sapiens
US-10-094-749-2039

Query Match          79.8%; Score 1105; DB 15; Length 364;
Best Local Similarity 98.6%; Pred. No. 3.4e-109;
Matches 214; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MEAHLTIRLTNKKVEFPFLVLLISGHCCLALVOGVSDFLLGKSLDIAPGMDLKVARR 60
Db 148 MEAHLTIRLTNKKVEFPFLVLLISGHCCLALVOGVSDFLLGKSLDIAPGMDLKVARR 207
```

```

CURRENT FILING DATE: 2005-04-04
PRIORITY APPLICATION NUMBER: 60/157,832
PRIORITY FILING DATE: 1999-10-05
PRIORITY APPLICATION NUMBER: 60/160,191
PRIORITY FILING DATE: 1999-10-19
PRIORITY APPLICATION NUMBER: 60/161,932
PRIORITY FILING DATE: 1999-10-28
PRIORITY APPLICATION NUMBER: 60/164,769
PRIORITY FILING DATE: 1999-11-12
PRIORITY APPLICATION NUMBER: 60/173,383
PRIORITY FILING DATE: 1999-12-28
PRIORITY APPLICATION NUMBER: 60/175,693
PRIORITY FILING DATE: 2000-01-12
PRIORITY APPLICATION NUMBER: 60/184,831
PRIORITY FILING DATE: 2000-02-24
PRIORITY APPLICATION NUMBER: 60/191,637
PRIORITY FILING DATE: 2000-03-23
NUMBER OF SEQ ID NOS: 43008
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 34191
LENGTH: 409
TYPE: PRT
ORGANISM: DROSOPHILA
US-11-097-143-34191

Query Match          30.0%; Score 415.5; DB 20; Length 409;
Best Local Similarity 38.2%; Pred. No. 2.7e-35;
Matches 100; Conservative 43; Mismatches 104; Indels 15; Gaps 6;

QY      1 MEAAATIRLTN--KYAFPEFLVLLISGGHCLALVQGVSDFLKSLDIAFGMDLQVA 58
        ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
DB      136 MEHAIQANNEHEQIQYFPLCLLASGHCQLVAVANGRLTLTGQLDAPGAPFKIG 195
        ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :

QY      59 RRLSLIKHEPCSTMSGKAIETILAK--QGNRFHFDIKPEPLHAANKDPSFGLQHYTKII 117
        ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
DB      196 RRLRLILPEYRLNMNGRAIEHAAQLASDPLAVEFPLPLAQORNCNPSFAGIKKNSFRAI 255
        ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :

QY      118 MKKEKEGIEKQIILSSADIAATVCHTMACHIVKTRHRAILFC--KQPDLLPENNAVLV 175
        ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
DB      256 RAERERARRPPDGVISNYGDFCAGLRSVSRHLMERTQRAIEYCLLPHRQLFGDTPPLTV 315
        ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :

QY      176 ASGVASNFYIRALLETITATQCTLLCPPEPLCTDNGIMIANWGIERLPAGLIHDIE 235
        ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
DB      316 MSGGVANNDAIVANIEHLLAAQYGCRRSPRSKRYCSDNGVMIAMHGVQL-----LDKKE 369
        ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :

QY      236 -GIRYEPKCPGLVDISKVEGEA 256
        ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
DB      370 ASTRYDYD--SIDIQSAGFA 388
        ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :

RESULT 15
US-10-067-443-3
/ Sequence 3, Application US/10067443
/ Publication No. US20030082782A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
FILE REFERENCE: D0073 NP
CURRENT APPLICATION NUMBER: US/10/067,443
CURRENT FILING DATE: 2002-02-05
PRIORITY APPLICATION NUMBER: US 60/266,518
PRIORITY FILING DATE: 2001-02-05
PRIORITY APPLICATION NUMBER: US 60/282,814
PRIORITY FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 463
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-10-067-443-3

```

Query Match 29.6%; Score 409.5; DB 14; Length 463;  
 Best Local Similarity 36.6%; Pred. No. 1.4e-34;  
 Matches 100; Conservative 37; Mismatches 91; Indels 45; Gaps 6;

QY	1	MEAHALTRILT-NKVEPFLVLLISGGHCLALVQVSDPFLLGKSLDIAPGMDLKVVAR	59
DB	192	MEAHALVARLVEOELSPFMALLISGHNLLVLAHKQYQLGTTVDDAIGEAFFDKTK	251
QY	60	RLSLIKHPECSTWSGKAIEHLAKQGNRFHFDIKRPLHAAKNCDFSTGLQHTDKIIMK	119
DB	252	WLGIDMH-----RSGPVEEELALEGDAKSVKFNVPKMKYKDCNPSYAGLKTQVRLAIEA	306
QY	120	KEKEGIEKGQIILSSADIATVQHTMACHLVKRTHRAILFCQRDLLPONNAVIVASGG	179
DB	307	KE-----INNRAIDIASFORVAVLHLEKCCERAIIDWALE--LEPSIKHNVISGG	353
QY	180	VASNFYIRRALEILTNAQTCTLCPPRILCTDNGIMIAMNGIERLRAGIGLHDIEGIRY	239
DB	354	VASNKYVRLRLNIVENKNLKLVCPPSLCTDNGVVMVMTGLEHFRVG-----RY	403
QY	240	E-----PKCPGLGVDISKEVGEA	256
DB	404	DPPPEATEPEDVYVDLRPRWPLGEEYAKGRSEA	436

Search completed: November 10, 2005, 16:58:51  
 Job time : 71.5228 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:23:22 ; Search time 16.5424 Seconds  
(without alignments)  
1552.972 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414  
 Report: 1295

Sequence: 1 MEAHALTIRLTNKVEFPFLV.....DISKEVEGASIKVPQLKMEI 267

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

```
Database :
1: pir_79:*
2: pir2:*
3: pir3:*
4: -pir4:*
```

**Pred.** No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	409.5	29.6	463	2	E84888	probable O-sialoglycoprotein
2	338	24.4	365	2	AB2902	O-sialoglycoprotein
3	338	24.4	366	2	D97677	probable O-sialoglycoprotein
4	338	24.4	387	2	E71711	probable O-sialoglycoprotein
5	334.5	24.2	359	2	AB3274	O-sialoglycoprotein
6	329.5	23.8	344	2	E97707	O-sialoglycoprotein
7	326.5	23.6	367	2	F87257	peptidase M22 family
8	291.5	21.0	341	2	H83527	O-sialoglycoprotein
9	290.5	21.0	335	2	G70369	O-sialoglycoprotein
10	289.5	20.9	342	2	H64079	O-sialoglycoprotein
11	279.5	20.2	337	2	A10079	probable glycoprotein
12	279.5	20.2	337	2	AG0892	probable glycoprotein
13	278.5	20.1	337	2	C91122	probable O-sialoglycoprotein
14	278.5	20.1	337	2	B85677	probable O-sialoglycoprotein
15	276	19.9	421	2	T18825	hypothetical protein
16	275.5	19.9	337	1	Q0EKR6	O-sialoglycoprotein
17	275	19.9	412	2	T40899	probable proteinase
18	270.5	19.5	325	2	A38108	O-sialoglycoprotein
19	265.5	19.2	354	2	C81040	O-sialoglycoprotein
20	263.5	19.0	354	2	C81886	probable O-sialoglycoprotein
21	263	19.0	348	2	D82807	O-sialoglycoprotein
22	256	18.5	346	2	H70195	O-sialoglycoprotein
23	253	18.3	346	2	C97888	O-sialoglycoprotein
24	250.5	18.1	346	2	F69786	glycoprotein endoprotease
25	250	18.1	346	2	B95015	glycoprotein endoprotease
26	244.5	17.7	343	2	D63716	glycoprotein endoprotease
27	243	17.5	327	2	G72411	hypothetical protein
28	242.5	17.5	348	2	S75548	O-sialoglycoprotein
29	241.5	17.4	346	2	AF1820	O-sialoglycoprotein

30	239.5	17.3	344	2	AC1334	glycoprotein endop
31	237.5	17.1	340	2	B97011	probably O-sialogl
32	236.5	17.1	336	2	B84936	O-sialoglycoprotei
33	233	16.8	323	2	G69388	O-sialoglycoprotei
34	231	16.7	344	2	AB1705	glycoprotein endop
35	229.5	16.6	346	2	G86651	O-sialoglycoprotei
36	222.5	16.1	324	2	F75029	O-sialoglycoprotei
37	220	15.9	338	2	A71545	probably O-sialogl
38	219.5	15.8	534	2	H69056	O-sialoglycoprotei
39	219	15.7	335	2	B81278	probably glycoprote
40	218	15.7	344	2	H72106	O-sialoglycoprotei
41	218	15.7	344	2	B86515	O-sialoglycoprotei
42	217	15.7	341	2	G89996	hypothetical prote
43	216	15.6	407	2	SS0740	ORF protein - yea
44	215.5	15.6	324	2	C71215	O-sialoglycoprotei
45	215.5	15.6	344	2	H70737	probable O-sialogl

## ALIGNMENTS

## RESULT 1

Probable O-rialdoglycoprotein endopeptidase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Jun-2003  
C:Accession: E84888  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vankken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; PMID:20083487; PMID:10617157

Query Match	29.6%	Score 409.5;	DB 2;	Length 463;
Best Local Similarity	36.6%;	Pred. No. 5.1e-29;		
Matches 100;	Conservative 37;	Mismatches 91;	Indels 45;	Gaps 6;

  

Oy	1	MEAAHLLTRILT-NKVEFPFLVLLISGGHCCLATVQGVSDFFLLGSLDIAPEMDLKVAR	59
Dd	192	MEAAHLVARLVSGEOLSFPMALLISGHNLVLVLAHKIQGYQLGTVDALGEADPKTKA	251
Oy	60	RLSLIKHPECSTMSGKAI EHLAKOGRFPHDIKPRLHAAXNCDFSFTGLDHVTDKIWK	119
Dd	252	WLGIDMH-----RSGGPVBEALLEGDAKSIVFNVPMKYHKDCNFSVAGLKQVRLAIEA	306
Oy	120	KEKEGIEKGQLSSAADTAATVQHTMACHLVKRTHEAILFCCKORDLLPONNAVLYASGG	179
Dd	307	KE-----IRNRDAIASFORVAVHLBEECCERAI DIALR--LEPSIKHWVISGG	353
Oy	180	VASNFIYIRALEILTNATQCTLCPPRLCTDNGIMTAMNGIERLRAGIGIHIDEGRFY	239
Dd	354	VASNKTIVRLNNIVENKNLIKVCPPSLCTDNGVMMAWTLEHFRVG-----RY	403
Oy	240	E-----PKCPLGVDISKVEGEA	256
Dd	404	DPPPATREPDYVDLRFPMPLGEAYAKGRSEA	436

RESULT 2  
AB2902  
O-sialoglycoprotein endopeptidase gcp [imported] - Agrobacterium tumefaciens (strain C58)  
C:\Species: Agrobacterium tumefaciens  
Date: 11-Jan-2002 #sequence revision 11-Jan-2002 #text change 18-Nov-2002

C:\Accession: AB2902  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, C.D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCell  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krepsan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.

A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens Cs8.  
A:Reference number: AB2577; PMID:21608550; PMID:11743193

A:Accession: AB2902  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-365 <KUR>  
A:Cross-references: GB:AEO08688; PIDN:ALA43632.1; PID:g17741154; GSPDB:GN00186  
A:Experimental source: strain CS8 (Dupont)  
C:Genetics:  
A:Gene: gcp  
A:Map position: circular chromosome  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match            24.4%; Score 338; DB 2; Length 365;  
Best Local Similarity   34.9%; Pred. No. 1.1e-22;  
Matches         90; Conservative   40; Mismatches   98; Indels   30; Gaps   7;

OY          1 MEHAATITLTKNVPEPFVILLISGHCCLALVGVSDPFLLGKSIDIAFGMDLDKYARR 60  
            :|||||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db          LEGHALTARTTDGLSFYMLLVSGGHTQLVLVRVGSEBRWGTTIDDLAGAFPDTAKI 179  
            |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
OY          61 LSILIKHPECSWTMGSKAIEMHLAKOGNRFPDIPKLPHLAKNCDFSTGL----QHYTDKI 116  
            ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db          180 LGI-PYP-----GGRAVENAAAGDPDRFPLPRMNGEARLDFFSSGLKTAVROATAI 232  
            ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
OY          117 IMKKKEEGIEKGQIISSAADIAATVOHTMACHLVKRTIRAILFCOKRDLLPONNA--VL 174  
            ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db          233 APLESD-----IADICASFQKAVASRTLKDRIGRGLAREKVE--PPHINEPAL 279  
            ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
OY          175 VASGSVAASNYTRALEILTNTAOCTLLCPPRLCTDNGIMANWNIESTLRAGLTIHDI 234  
            |:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db          280 VVAGSVAANOETIKOTLOALCDTHGFRRFAVPHRILCTDNAAMTAWGLEMAEG---RQA 335  
            |:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
OY          235 EGIRYPEKCPLGDYSKE 252  
            ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db          336 DALEVAPRSRWPIDGSAE 353

RESULT 3  
D97677  
Probable o-sialoglycoprotein endopeptidase (glycoproteinas) [imported] - Agrobacterium  
C:Species: Agrobacterium tumefaciens  
C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C:Accession: D97677  
R:Goodner, F.; Hinkle, G.; Gatcung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldman,  
R.; Liu, F.; Wolam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.,  
Science 294, 2323-2328, 2001  
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tun  
A:Reference number: AY9359; PMID:21608551; PMID:11743194  
A:Accession: D97677  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-366 <KUR>  
A:Cross-references: GB:AEO07869; PIDN:AAK68373.1; PID:g15157858; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C 4806  
A:Map position: circular chromosome  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match            24.4%; Score 338; DB 2; Length 366;  
Best Local Similarity   34.9%; Pred. No. 1.1e-22;  
Matches         90; Conservative   40; Mismatches   98; Indels   30; Gaps   7;

OY          1 MEHAATITLTKNVPEPFVILLISGHCCLALVGVSDPFLLGKSIDIAFGMDLDKYARR 60  
            :|||||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db          LEGHALTARTTDGLSFYMLLVSGGHTQLVLVRVGSEBRWGTTIDDLAGAFPDTAKI 180  
            |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Oy 1 LSLIHGPEOSTMSGGATIEHLAKOGRPFHDIPRLHAKNCDESPFQI-----GHVTDKI 116  
 181 LGL-FYP-----GGPAVENAAAKGDDPRFPLRPFWGGEARLDSPFSGLKTAIVRQDAATRI 233  
 Oy 117 IMKKEKEGIEKGQILSSAADIAATVQHTMACHLVKTRTHAILECKQRDLLPQNNAA--VL 174  
 Db 234 APLISGQ-----IADIASFQKAVSRFLTKDRIIGRLAPKVE--PPIINGEPAL 280  
 Oy 175 VASGGVANSFYIRRALEILTMTAQTCLCPRPRLCTNGIMIANNGIERLAPAGLGILNDI 234  
 Db 281 VVAGGVANNOEIRQTLQALCDTHGFRFVAPPHRLCTNNAMITAWAGLERMAG-----ROA 336  
 Oy 235 EGIRYEPKCPGLGVDSIKE 252  
 Db 337 DALBVAIPRSRWPLDGSAAE 354  
 RESULT 4  
 E71711  
 Probable O-sialoglycoprotein endopeptidase (gcp) RP037 - Rickettsia prowazekii  
 C:Species: Rickettsia prowazekii  
 C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 09-Jul-2004  
 C:Accession: E71711  
 R:Accession: S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U.  
 Nature 396, 133-140, 1998  
 A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
 A:Reference number: A71630; MID:J99039499; PMID:9823893  
 A:Accession: E71711  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Residues: 1-387 <AND>  
 A:Molecule type: DNA  
 A:Cross-references: UNIPROT:O9ZEAB; GB:AJ235270; GB:AJ235269; NID:g3860572; PIDD:CAAL450  
 A:Experimental source: strain Madrid E  
 C:Genetics:  
 A:Gene: gcp;  
 C:Superfamily: O-sialoglycoprotein endopeptidase  
 Query Match 24.4%; Score 338; DB 2; Length 387;  
 Best Local Similarity 32.5%; Pred. No. 1,1e-22;  
 Matches 87; Conservative 38; Mismatches 83; Indels 60; Gaps 6;  
 Oy 1 MEAAHATRLTNKVEFPFLVLISGGHCLLALVQGVSDFLLGKSLDIAPGMDLKVARR 60  
 Db 113 LEGHALTARLTQNISSYPVLLILASGGHCQFVAVLGLKYYKILGTTIDDAVGETFPDKVAK 172  
 Oy 61 LSLIHGPEOSTMSGGATIEHLAKOGRPFHDIPRLHAKNCDESPFQI-----GHVTDKI 116  
 Db 173 LNL-----SPGGEIERRAALQGNPHKYPFPPTIISNGCNMSFSGLKTAIVRFLTNL 225  
 Oy 120 KEKEKEIEKGQILSSAADIAATVQHTMACHLVKTRTHAILECKQ----- 163  
 Db 226 KEVNDV-----INDIAPSFQITIGALISSMGQDAIRLYKQILNDIYEDINHPKLN 277  
 Oy 164 ----RD-----LLPON-----NAVLVASGGVANSFYIRRALEILT 195  
 Db 278 LKSPRDBENMKPLECITRPKYRIHIONSYSRNLNDITVIAGGVAAANKYQIEILSDCTR 337  
 Oy 196 ATQCTLLCPRPRLCTNGIMIANNGIER 223  
 Db 338 PYGRIIAPPMLCTDNAMITAVAGLER 365  
 RESULT 5  
 AB3274  
 O-sialoglycoprotein endopeptidase (EC 3.4.24.57) [imported] - Brucella melitensis (strain  
 C:Species: Brucella melitensis  
 C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 15-Feb-2002  
 C:Accession: AB3274  
 R:Accession: V.G.; Kaparatul, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
 R. DelVecchio, A.G.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

A:Reference number: AD3252; PMID:11756688  
A:Accession: AB3274  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-359 <KUR>  
A:Cross-references: GB:AE008917; PIDN:AA151357.1; PID:g17982056; GSPDB:GN00190  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BME10175  
A:Map position: 1  
C:Superfamily: O-sialoglycoprotein endopeptidase  
C:Keywords: hydrolase, metalloproteinase

Query Match 24.2%; Score 334.5; DB 2; Length 359;  
Best Local Similarity 37.6%; Pred. No. 2.2e-22;  
Matches 86; Conservative 33; Mismatches 83; Indels 27; Gaps 6;

1 MEAHLTLITLNKVEPFLVLLISGGHCLALVGVSPFLLGKSLDIAFGMLDVARR 60  
Db LEGHALLTALTDLPFPYLLLVSGGHQWMLVKGIGYERLGTITDDALGEAFDIAKL 175

61 LSLIKHPECSWSGKAIEHLAKQGNRFHDIKPELHAKNCDFSPFGI-QHVTDK 115  
Db LGL-PYP-----GGPVRWALGGDQKRFALPRLPKGEARLDFSPSGLKTAVROTATEL 228

116 IIMKEKEGIEKGQILSSAADIAATVQHTMACHLVKTRTHAILECKQDLPQ-NNAVL 174  
Db VPLTQ-----DVTDICASFQAAVADLTSDVGRSLERFKTE--PPDCATPSL 274

175 VASGVAASFYIRRALEILTNAQTCLCPPEPRLCTDNGIMAMNGIER 223  
Db VVAGVVAANKTILRAALENLCTRHGFAPITAPPLNLCTDAAIMAMGAER 323

RESULT 6  
E97707  
O-sialoglycoprotein endopeptidase (EC 3.4.24.57) [imported] - Rickettsia conorii (strain  
C:Species: Rickettsia conorii  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 22-Oct-2001  
C:Accession: E97707  
R:Ogata, H.; Audic, S.; Rensse-Audiffren, P.; Fournier, P. E.; Barbe, V.; Samson, D.; Rd  
Science 293. 2093-2098, 2001  
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A:Reference number: A97700; MUID:21442074; PMID:11557893  
A:Accession: E97707  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-344 <KUR>  
A:Cross-references: GB:AE006914; PIDN:AA102559.1; PID:g15619097; GSPDB:GN00173  
C:Genetics:  
A:Gene: gcp  
C:Superfamily: O-sialoglycoprotein endopeptidase  
C:Keywords: hydrolase, metalloproteinase

Query Match 23.8%; Score 329.5; DB 2; Length 344;  
Best Local Similarity 36.0%; Pred. No. 5.8e-22;  
Matches 81; Conservative 40; Mismatches 85; Indels 19; Gaps 5;

1 MEAHLTLITLNKVEPFLVLLISGGHCLALVGVSPFLLGKSLDIAFGMLDVARR 60  
Db LEGHALLTALTDLPFPYLLLVSGGHQWMLVKGIGYERLGTITDDALGEAFDIAKL 172

61 LSLIKHPECSWSGKAIEHLAKQGNRFHDIKPELHAKNCDFSPFGI-QHVTDKIMK- 119  
Db LNL-----AFPGGPEIERAKLGDPHKFKFPPIINSGCNMSFSGIKTAVRTLIWLT 225

120 KEKEGIEKGQILSSAADIAATVQHTMACHLVKTRTHAILECKQ-RDLLPQNNAAVLVAG 178  
Db KEINPTV-----INDIAASFQITIGELISSKYODAIRAYEQITNPPDKN--IVAG 275

179 VASNFYIRRALEILTNAQTCLCPPEPRLCTDNGIMAMNGIER 223  
Db VVAANKTILKLISSCAKTYGRILYPIHILCTDAAIMAMVAGLER 320

RESULT 7

peptidase M22 family protein [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001

C:Accession: F87257

R:Nickman, W.C.; Feldbylum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Labb, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon, n. U.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: F87257

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-367 <STO>

A:Cross-references: GB:AE005673; NID:g13421168; PIDN:AAK22058.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0071

C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 23.6%; Score 326.5; DB 2; Length 367;

Best Local Similarity 36.3%; Pred. No. 1.2e-21;

Matches 89; Conservative 36; Mismatches 99; Indels 21; Gaps 6;

OY 1 MEAAHTLRLTKVKEPFLVLLISGGHCLALVGVSDFLILGSLDIAPEGMDLKVARR 60

DB 124 LEGHAVSARLADIAVPFLILLVSGGHCQLLEVSGVACCKRLGTTIIDDAAGEAFDKIAKS 183

OY 61 LSLIHPECSMTSGGKAIEHLAKQGRFHFDIKPELHAKKCDSPFGLQHTDKIKKK 120

DB 184 LGL-PYP-----GGALERDLAVGSDPTRYALPRALGRKDCDFSGLKTAAARIATFL 236

OY 121 EKEEGIEKQILSSAADIAATVQHTWACHLVKTRTHAILPEKQKDLIPQNNAVLVAASGV 180

DB 237 TTDD-----ARRDLAAGVQAAIARQLSERVDRAMKLYK--DSHPDELRFVVAAGV 285

OY 181 ASNFYIRALIELITNATQCTLLCPPRCLCTDNGIMAMNGIERLRAGIGILHDIIGIRYE 240

DB 286 AANGAVRALLLADCEKNGFSFAPPLAYCTDNAMAMIALAGERL--ALGIFDDIDATA-R 342

OY 241 PKCPL 245

DB 343 PRWPL 347

RESULT 8

H83572

O-sialoglycoprotein endopeptidase PA0580 [imported] - Pseudomonas aeruginosa (strain PAO

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: H83572

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Micozuchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lim, A.; Lory, S.; Olson, M.V. Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathob

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: H83572

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-341 <STO>

A:Cross-references: GB:AE004494; GB:AE004091; NID:g9946446; PIDN:AG03969.1; GSPDB:GN001

A:Experimental source: strain PAO1

C:Genetics:

A:Gene: gcp; PA0580

C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 21.0%; Score 291.5; DB 2; Length 341;

Best Local Similarity 35.4%; Pred. No. 1.5e-18;

Matches 87; Conservative 37; Mismatches 99; Indels 23; Gaps 9;



QY 1 MEAHALTRIRLTK-VEPPFLVLLISGGHCLALVQGVSDPFLLGKSLDIAPGMDLKVAR 59  
Db 112 MEGHLLAPMLEBQPRPFVALLVSGHTQVLVRVDGIRYQLLGSSVDDAAGEADKTXK 171  
QY 60 RLSLIKHEPCSTMSGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTVDKTIK 119  
Db 172 LIGL-GYP-----GGEPIATLARSGTPGRFVFPFMDRPGDLFSGLGKTFITLN-TWQ 223  
QY 120 KEKEBIEKGQILSSADIAATVQHTMACHLVKRTHRALIFCKQRDLPPONNAVIVASGG 179  
Db 224 RCVEAGDSEQ---TRCDIALAFDTAVVETLIIKRRRL---KQYGL--KN--LVVAGG 272  
QY 180 VASNFYIRRALIEITNATQCTLLCPRPRLCTDNGIMIAMNGIERLRAGLGILHDIEGIRY 239  
Db 273 VSNQALRSGLGEKMLGEMKQGVFARPRFCTDNGMIMIVACQRLLAG---QHDGPAISV 329  
QY 240 EPRKCP 245  
Db 330 QPRWPM 335

## RESULT 9

siatologlycoproteinase - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004  
C:Accession: G70369  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: G70369  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-335 <AOE>  
A:Cross-references: UNIPROT:O66986; GB:AB000708; NID:g2983356; PIDN:AAC06951.1; PID:g298  
A:Experimental source: strain VFS  
C:Genetics:  
A:Gene: gcp  
C:Superfamily: O-siatologlycoprotein endopeptidase

Query Match 21.0%; Score 290.5; DB 2; Length 335;  
Best Local Similarity 34.3%; Pred. No. 1.8e-18;  
Matches 85; Conservative 46; Mismatches 82; Indels 35; Gaps 10;  
QY 1 MEAHALTRIRLTKVEPPFLVLLISGGHCLALVQGVSDPFLLGKSLDIAPGMDLKVAR 60  
Db 112 LBGHIVSFLEKTYEPLALIIISGHTDIVLRDFGRYDFLGTLDDAVGEAYDKVAKM 171  
QY 61 LSLIKHEPCSTMSGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTVDKTIK 120  
Db 172 LIGL-GYP-----GGEPIATLARSGKCL-YPLPKLMEEGNLNFSGLK--TAILML 220  
QY 121 EKEBIEKGQILSSADIAATVQHTMACHLVKRTHRALIFCKQRDLPPONNAVIVASGG 180  
Db 221 KKEKNVRK-----EDLAYSFOETVAILLEKS---LWAKKTKGIR---LVVAGV 265  
QY 181 ASNFYIRRALIEITNATQCTLLCPRPRLCTDNGIMIAMNGIERLRAGLGILHDIEGI 237  
Db 266 SANSRLR---EVFKASQYGFELYIPIPSLSTDAIMAIYAGMERFKGVAPLDVNP- 321  
QY 238 RYEPKCP 245  
Db 322 --QPNIFL 327

## RESULT 10

H64074  
O-siatologlycoprotein endopeptidase (BC 3.4.24.57) - Haemophilus influenzae (strain Rd KW2  
N:Alternate names: siatologlycoproteinase  
C:Species: Haemophilus influenzae

C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 09-Jul-2004  
C:Accession: H64074  
R:Feischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kjerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.  
Science 269, 436-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: H64074  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-342 <TIGR>  
A:Cross-references: UNIPROT:P43764; GB:U32735; GB:I42023; NID:g1573509; PIDN:AAC22187.1;  
C:Superfamily: O-siatologlycoprotein endopeptidase  
C:Keywords: hydrolase; metalloproteinase

Query Match 20.9%; Score 289.5; DB 2; Length 342;  
Best Local Similarity 35.7%; Pred. No. 2.3e-18;  
Matches 82; Conservative 30; Mismatches 95; Indels 23; Gaps 7;

QY 1 MEAHALTRIRL-TNKVEPPFLVLLISGGHCLALVQGVSDPFLLGKSLDIAPGMDLKVAR 59  
Db 112 MEGHLLAPMLDNPSPHFPFVALLVSGHTQVLVRVDGKYEVLGESSIDPAAGEAFDKTK 171  
QY 60 RLSLIKHEPCSTMSGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTVDKTIK 119  
Db 172 LIGL-DYP-----GGAALSRLEKGTTPRFPFRPMTDRAGIDFSGLGKTFPAANTVNG 224  
QY 120 KEKEBIEKGQILSSADIAATVQHTMACHLVKRTHRALIFCKQRDLPPONNAVIVAS 177  
Db 225 AIKNGELIEQ-----TKADIAVAFQDAVVDILA-----IKCK-RALKETGYKRLVIA 271  
QY 178 GGVASNFYIRRALIEITNATQCTLLCPRPRLCTDNGIMIAMNGIERLRAG 227  
Db 272 GGVASANKKLRITLAHLMONIGEVFPPOPFCFTDNGAMIAATYGLRLKOG 321

## RESULT 11

AI0079  
probable glycoproteinase gcp [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
C:Accession: AI0079  
R:Parhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.  
; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
11, M.; Rutherford, K.; Simmonds, W.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AI0079  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-337 <KUR>  
A:Cross-references: GB:AL590442; PIDN:CAC89500.1; PID:g15978736; GSPDB:GN00175  
C:Genetics:  
A:Gene: gcp  
C:Superfamily: O-siatologlycoprotein endopeptidase

Query Match 20.2%; Score 279.5; DB 2; Length 337;  
Best Local Similarity 32.8%; Pred. No. 1.8e-17;  
Matches 81; Conservative 35; Mismatches 102; Indels 29; Gaps 7;

QY 1 MEAHALTRIRL-TNKVEPPFLVLLISGGHCLALVQGVSDPFLLGKSLDIAPGMDLKVAR 59  
Db 112 MEGHLLAPMLEBAPFPFVALLVSGHTQILSVTIGEYLLIGSVDDAAGEAFDKTK 171  
QY 60 RLSLIKHEPCSTMSGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTVDKTIK 119  
Db 172 LIGL-DYP-----GGEPIATLARSGKCL-YPLPKLMEEGNLNFSGLK--TAILML 224  
QY 120 KEKEBIEKGQILSSADIAATVQHTMACHLVKRTHRALIFCKQRDLPPONNAVIVASG 178  
Db 225 KKEKNVRK-----EDLAYSFOETVAILLEKS---LWAKKTKGIR---LVVAGV 265



QY	240	EPKCPL	245
		:	
Db	326	RPRWPL	331

### RESULT 15

hypothesized Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000  
 C:Accession: T18825  
 R:Matthews, L.  
 submitted to the EMBL Data Library, October 1996  
 A:Reference number: Z19027  
 A:Accession: T18825  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-421 <Wt>  
 A:Cross-references: EMBL: Z81030; PIDN: CAB02716.1; GSPDB: GNM00023; CESP: C01G10.10  
 A:Experimental source: clone C01G10  
 C:Genetics:  
 A:Gene: CESP:C01G10.10  
 A:Map position: 5  
 A:Introns: 51/2, 72/3, 122/2, 177/1, 272/3, 315/1, 353/2  
 A:Superfamily: O-staloglycoprotein endopeptidase

Query Match 19.9%; Score 276; DB 2; Length 421;  
Best Local Similarity 30.4%; Pred. No. 5e-17;  
Matches 85; Conservative 52; Mismatches 107; Indels 36; Gaps 8;

```

Oy      1 MEAAHATRLT-NKVEFPVLLISGHCCLMAVOVSPELLGSLDIAFGMDLVAR 59
Db      134 MRAHASTILVDSVRFPFSVAHLISGCHALISVAEDEVFKLYGVSQSGPFCIDIKVAR 1933
Oy      60 RLSLIKHPECSTMSGKALEHLAKOQN---RFHFIDIKPLHNAKCNDFSTGLQHVTDKI 116
Db      194 QLGDE-GSEFFDGIHVGAAVEIILASRASAQGHLYRPIFLPNVPKANNFPODKISYYINLE 252
Oy      117 IMKKEKEBEGIEKGQILSSAADIAATQYOHMACHLVKTRHALLFCOKRDLDPQNNATLVA 176
Db      253 RLKRSKSESTD-----IPDFASLQNTVAARIHSKHLFFESLSBEQELKPO--LVI 302
Oy      177 SGGVANFYIRARALEILTNATOCTLLCPPRLCCTDNGMIAMNGIERLRAGLGIHLHDIEG 236
Db      303 GGGVANQYIRFGAISKLSAAHNVTITIKVLLSCTDINAEMIAEYSGL-----LMLVNRSEA 356
Oy      237 IRYEP-----KCPGLVDISKEVGEASIKVQLKM 265
Db      357 IWMRPNDIPDTYAAHARSIDIGTASSEI---LDTPRKX 392

```

Search completed: November 10, 2005, 16:53:08  
Job time : 17.5424 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: November 10, 2005, 15:16:01 ; Search time 64.7185 Seconds  
(without alignments)  
2112.614 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414  
Perfect score: 1385  
Sequence: 1 MESHALLTRITNKVPEPFLV.....DISKEVGASIKVQLKREI 267

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_03: \*  
1: uniProt\_sprot: \*  
2: uniProt\_trembl: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1385	100.0	414	2	096EV9
2	1358.5	98.1	439	2	09H4B0
3	1231	88.9	414	2	06PEB4
4	1224	88.4	414	2	08BLB6
5	1216	87.8	414	2	09D0N0
6	1116	80.6	467	2	06AYN7
7	1105	79.8	364	2	096NH5
8	859	62.0	404	2	08JFW3
9	855	61.7	404	2	08JFR7
10	453.5	32.7	401	2	0709I8
11	415.5	30.0	323	2	09V0S6
12	415.5	30.0	409	2	09W0D6
13	407.5	29.4	480	2	022145
14	369	26.6	335	2	073H71
15	356.5	25.7	360	2	092LH8
16	352	25.4	362	2	098ET6
17	349	25.2	387	2	068XR3
18	342	24.7	251	2	093FU2
19	339.5	24.5	389	2	06ND54
20	338	24.4	365	2	08UC47
21	338	24.4	366	2	07CM48
22	338	24.4	387	1	GCP_RICPR
23	334.5	24.2	359	2	08YUJ1
24	334.5	24.2	359	2	08FYI5
25	333.5	24.1	344	2	07PAG7
26	329.5	23.8	344	2	092JX6
27	326.5	23.6	357	2	089MW1
28	326.5	23.6	357	2	G5A25
29	321.5	23.2	346	2	07VXN4
30	320	23.1	364	2	06G1R3
31	317	22.9	340	2	06FCR9

32	316.5	22.9	346	2	07M668	07M668 bordetella
33	315.5	22.8	346	2	07W134	07W134 bordetella
34	304.5	22.0	343	2	09CLJ1	09CLJ1 pasteurella
35	304	21.9	353	2	07VQO9	07VQO9 candidatus
36	302.5	21.8	341	2	07NUE3	07NUE3 chromobacter
37	302	21.8	339	2	06LV10	06LV10 photobacter
38	299.5	21.6	255	2	0677H2	0677H2 hyacinthus
39	298	21.5	364	2	06FYF1	06FYF1 bartonella
40	293	21.2	344	2	06SRP0	06SRP0 pseudomonas
41	291.5	21.0	341	2	091SV7	091SV7 pseudomonas
42	290.5	21.0	335	1	GCP_AOUAE	06986 aquifex aeo
43	289.5	20.9	342	1	GCP_HAEIN	043764 haemophilus
44	286.5	20.7	337	2	082XN2	082XN2 nitrosomona
45	285.5	20.6	348	2	09L7A5	09L7A5 haemophilus

## ALIGNMENTS

RESULT 1  
096EV9 PRELIMINARY: PRT; 414 AA.  
AC 096EV9  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE OSCEPL1 protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Muscle;  
RA MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strassburg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schaller G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schaefer T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.N., Sodergren B.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,  
RA Jones S.J., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences".  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Muscle;  
RA Strassburg R.L.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC011904; AAH11904.1; -  
DR MEROPS; M22.004; -  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR00905; Peptidase M22.  
DR InterPro: IPR00180; Pept M22\_Osialgl.  
DR Pfam: PF00814; Peptidase M22\_1.  
DR PIRSF; PIRSF004537; Osialglc\_Optds; 1.  
DR PRINTS; PR00789; Osialglc\_Optds; 1.  
DR Prodom; PD002367; Peptidase M22; 1.  
DR TIGFAMS; TIGR00329; GCP; 1.  
DR SEQUENCE 414 AA; 45122 MW; A536B333F5C6B8DD CRC64;

Query Match 100.0%; Score 1385; DB 2; Length 414;  
Best Local Similarity 100.0%; Pred. No. 2,9e-110;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAHALTRITLNKVEPFLVLLISGHCILALVQVSDPFLILGKSLDIAFGMDLKVARR 60  
DB 148 MEAHALTRITLNKVEPFLVLLISGHCILALVQVSDPFLILGKSLDIAFGMDLKVARR 207  
QY 61 LSLIKHEPCSTMSGKAIENLAKOGRNRFHDIKPLHAKNCDSFTGLQHTYDKIIMKK 120  
DB 208 LSLIKHEPCSTMSGKAIENLAKOGRNRFHDIKPLHAKNCDSFTGLQHTYDKIIMKK 267

QY 121 EKEGIEKGQILSSADIAATVQHTMACHLVKTTHRALIFCKQRDLTPONNAVIVASGV 180  
DB 268 EKEGIEKGQILSSADIAATVQHTMACHLVKTTHRALIFCKQRDLTPONNAVIVASGV 327

QY 181 ASNFYIRRALEILTNATQCTLLCPPRLCTDNGMIANNGIERLRAGILHIDIGIRYE 240  
DB 328 ASNFYIRRALEILTNATQCTLLCPPRLCTDNGMIANNGIERLRAGILHIDIGIRYE 387

QY 241 PKCPGLVDISKVEGASIKVPLKMEI 267  
DB 388 PKCPGLVDISKVEGASIKVPLKMEI 414

RESULT 2  
Q9H4B0 PRELIMINARY; PRT; 439 AA.  
ID Q9H4B0  
AC Q9H4B0; 2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)  
DE Putative sialoglycoprotein type 2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Uterus;  
RA Chen J.M., Fortunato M., Barrett A.J.;  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ295148; CAC14666.1; -  
DR MEROPS: M22.004; -  
DR Genew: HENC:23075; OSGEPL1.  
DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO: GO:0008233; F:peptidase activity; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR000905; Peptidase\_M22.  
DR InterPro: IPR009180; Pept\_M22\_Osialgl.  
DR Pfam: PF00814; Peptidase\_M22; 1.  
DR PIRSF: PIRSF004537; Osialglc\_ptds; 1.  
DR PRINTS: PR00789; OSIALOPTASE.  
DR Prodom: PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs: TIGR00329; gcp; 1.  
KW Protease.  
SQ SEQUENCE 439 AA; 48040 MW; 44849372C784E41F CRC64;

Query Match 98.1%; Score 1358.5; DB 2; Length 439;  
Best Local Similarity 91.1%; Pred. No. 5.8e-106;  
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

QY 1 MEAHALTRITLNKVEPFLVLLISGHCILALVQVSDPFLILGKSLDIAFGMDLKVARR 60  
DB 148 MEAHALTRITLNKVEPFLVLLISGHCILALVQVSDPFLILGKSLDIAFGMDLKVARR 207  
QY 61 LSLIKHEPCSTMSGKAIENLAKOGRNRFHDIKPLHAKNCDSFTGLQHTYDKIIMKK 120  
DB 208 LSLIKHEPCSTMSGKAIENLAKOGRNRFHDIKPLHAKNCDSFTGLQHTYDKIIMKK 267  
QY 121 EKEGI-----EKGQILSSADIAATVQHTMACHLVKTTH 155

DB 268 EKEGIEKIFISKVEQINIPGLCIKAHFCRYEKQILSSADIAATVQHTMACHLVKTTH 327  
QY 156 RALIFCKQRDLTPONNAVIVASGVASNFYIRRALEILTNATQCTLLCPPRLCTDNGMI 215  
DB 328 RALIFCKQRDLTPONNAVIVASGVASNFYIRRALEILTNATQCTLLCPPRLCTDNGMI 387

QY 216 IAMNGIERLRAGILHIDIGIRYEPKCPGLVDISKVEGASIKVPLKMEI 267  
DB 388 IAMNGIERLRAGILHIDIGIRYEPKCPGLVDISKVEGASIKVPLKMEI 439

RESULT 3  
Q6PEB4 PRELIMINARY; PRT; 414 AA.  
ID Q6PEB4  
AC Q6PEB4; 2004 (TREMblrel. 27, Created)  
DT 05-JUL-2004 (TREMblrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMblrel. 27, Last annotation update)  
DE Hypothetical protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CZECH II; TISSUE=Mammary tumor;  
RX MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smallos D.E., Scherch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CZECH II; TISSUE=Mammary tumor;  
RA Strausberg R., Feingold E.A., Grouse L.H., Derge J.G.,  
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC058172; AAH58172.1; -  
DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR000905; Peptidase\_M22.  
DR InterPro: IPR009180; Pept\_M22\_Osialgl.  
DR Pfam: PF00814; Peptidase\_M22; 1.  
DR PIRSF: PIRSF004537; Osialglc\_ptds; 1.  
DR PRINTS: PR00789; OSIALOPTASE.  
DR Prodom: PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs: TIGR00329; gcp; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 414 AA; 44962 MW; 0418AD93C6FEBCAE CRC64;

Query Match 88.9%; Score 1231; DB 2; Length 414;  
Best Local Similarity 87.3%; Pred. No. 4.6e-97;  
Matches 233; Conservative 15; Mismatches 19; Indels 0; Gaps 0;

QY 1 MEAHALTRITLNKVEPFLVLLISGHCILALVQVSDPFLILGKSLDIAFGMDLKVARR 60

Db 148 MEAHALTRLRNKNKFFPLVLLISGHCILALVGVSDPFLILGKSLDIAPGMDLKVARR 207  
Qy 61 LSLIKHEPCSTMSGKAIIEHLAKQGNRPHFDIKPPLHAKKCDPFTGLQHVTDKIIMKK 120  
Db 208 LSLIKHEPCSTMSGKAIIEHLAKQGNRPHFTINPMQNAKCDPFTGLQHVTDKILTHK 267  
Qy 121 EKEBIEKGQILSSAADIAAVQHTACHLAKRTHRALFCQKQNLSPANAVLVASGV 180  
Db 268 EKEBIEKGQILSSAADIAAVQHTACHLAKRTHRALFCQKQNLSPANAVLVASGV 327  
Qy 181 ASNFYIRALREILTNATQCTLLCPPRLCTDNGIMIANWGIERLPAAGILHDIGIRYE 240  
Db 328 ASNLTYIRKALREIVANATQCTLLCPPRLCTDNGIMIANWGIERLPAAGILHDIGIRYE 387  
Qy 241 PKCPLGVDISKEVGEASIKVPOLKMEI 267  
Db 388 PKCPLGVDISREVAEAAIKVRLKVAL 414  
RESULT 4  
Q8BLB6 PRELIMINARY; PRT; 414 AA.  
ID Q8BLB6  
AC 08BLB6  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length  
DE enriched library, clone:B30219017 product:similar to PUTATIVE  
DE SIALOGLYCOPROTEASE TYPE 2.  
GN Name=Osgp11;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Meth. Enzymol. 303:19-44(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA RIKEN FANTOM Consortium;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagakura S., Sasaki N., Carninci P.,  
RA Kono H., Akiyama J., Nishi K., Kitamura T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,

RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
RA Hayashida K., Hayatsu N., Hiramoto K., Hirooka T., Hirozane T.,  
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,  
RA Kurihara C., Matsumura T., Miyazaki A., Murata M., Nakamura M.,  
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
RA Saito R., Satoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AK045669; BAC32450.1; -.  
DR MEROPS: M22.004; -.  
DR MGI: 1919335; Osgp11.  
DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO: GO:0008233; F:peptidase activity; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR000905; Peptidase\_M22.  
DR InterPro: IPR009180; Pept\_M22\_Osialg1.  
DR Pfam: PF00814; Peptidase\_M22; 1.  
DR PIRSF: PIRSF004537; Osialgic\_ptlde; 1.  
DR PRINTS: PR00789; OSIALOPTASE.  
DR Prodom: PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs: TIGR00329; gcp; 1.  
KM Protease.  
SQ SEQUENCE 414 AA; 44927 MW; 9A2D27B576C7DB24 CRC64;  
Query Match 88.4%; Score 1224; DB 2; Length 414;  
Best Local Similarity 87.3%; Pred. No. 1.8e-96;  
Matches 233; Conservative 14; Mismatches 20; Indels 0; Gaps 0;  
Qy 1 MEAHALTRLRNKNKFFPLVLLISGHCILALVGVSDPFLILGKSLDIAPGMDLKVARR 60  
Db 148 MEAHALTRLRNKNKFFPLVLLISGHCILALVGVSDPFLILGKSLDIAPGMDLKVARR 207  
Qy 61 LSLIKHEPCSTMSGKAIIEHLAKQGNRPHFDIKPPLHAKKCDPFTGLQHVTDKIIMKK 120  
Db 208 LSLIKHEPCSTMSGKAIIEHLAKQGNRPHFTINPMQNAKCDPFTGLQHVTDKILTHK 267  
Qy 121 EKEBIEKGQILSSAADIAAVQHTACHLAKRTHRALFCQKQNLSPANAVLVASGV 180  
Db 268 EKEBIEKGQILSSAADIAAVQHTACHLAKRTHRALFCQKQNLSPANAVLVASGV 327  
Qy 181 ASNFYIRALREILTNATQCTLLCPPRLCTDNGIMIANWGIERLPAAGILHDIGIRYE 240  
Db 328 ASNLTYIRKALREIVANATQCTLLCPPRLCTDNGIMIANWGIERLPAAGILHDIGIRYE 387  
Qy 241 PKCPLGVDISKEVGEASIKVPOLKMEI 267  
Db 388 PKCPLGVDISREVAEAAIKVRLKVAL 414  
RESULT 5  
Q9DON0 PRELIMINARY; PRT; 414 AA.  
ID Q9DON0  
AC Q9DON0  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DE Mus musculus 10 days embryo whole body cDNA, RIKEN full-length  
DE enriched library, clone:2610001m19 product:similar to PUTATIVE  
DE SIALOGLYCOPROTEASE TYPE 2.

GN Name=Osegepl1;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=whole body;  
RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Meth. Enzymol. 303:19-44(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=whole body;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055560;  
RA RIKEN FANTOM Consortium;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=whole body;  
RA The FANTOM Consortium;  
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=whole body;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=whole body;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
Sami N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kasaiwa K.,  
Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Wachihi K.,  
Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=whole body;  
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
Imotani K., Ishii Y., Itoh M., Izawa M., Kakuwa T., Kato H.,  
Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,  
Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
Yuramatsu M., Hayashizaki Y.;  
RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AK011265; BAB37506.1; -;  
DR MEROPS; M22.004; -;  
DR MGD; MGI:1919335; Osegepl1.  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis; P:peptidolysis; IEA.  
DR Interpro; IPR00905; Peptidase\_M22.

DR Interpro; IPR009180; Pept\_M22\_Osialgl.  
DR Pfam; PF00814; Peptidase\_M22; 1.  
DR PIRSF; PIRSF004537; Osialglc\_ptids; 1.  
DR PRINTS; PR00789; OSIALOPTASE.  
DR Prodom; PD002367; Peptidase\_M22; 1.  
DR TIGRfam; TIGR00329; gcp; 1.  
KM Peptidase.  
SQ SEQUENCE 414 AA; 44999 MW; 999BC689944DB24 CRC64;  
Query Match 87.8%; Score 1216; DB 2; Length 414;  
Best Local Similarity 86.9%; Pred. No. 8.9e-96;  
Matches 232; Conservative 14; Mismatches 21; Indels 0; Gaps 0;  
QY 1 MEAHLITRLTNKVEFPPLVILLISGHCILLAVGVSPFLIGKSLDIAPGMDLVARR 60  
DB 148 MEAHLITRLTNKVEFPPLVILLISGHCILLAVGVSPFLIGKSLDIAPGMDLVARR 207  
QY 61 LSLIHPECSITMSGKALIEHLAKGNRPFDIKPPLHAKNCDPFTGQHTDKIMKK 120  
DB 208 LSLIHPECSITMSGKALIEHLAKGNRPFDIKPPLHAKNCDPFTGQHTDKIMKK 267  
QY 121 EKEEGIEKGQILSSAADIAATVQHTMACHLVKTTRAILFCRKORDLPONNAVIVASGV 180  
DB 268 EKEEGIEKGQILSSAADIAATVQHTMACHLVKTTRAILFCRKORDLPONNAVIVASGV 327  
QY 181 ASNPYIRRALETITLNTAOTCTLLCPPRLCTDNGIMIANWGIERRAGILHDIGIRYE 240  
DB 328 ASNPYIRRALETITLNTAOTCTLLCPPRLCTDNGIMIANWGIERRAGILHDIGIRYE 387  
QY 241 PKCPPLGVDSIKVEGASIKVPOKMEI 267  
DB 388 PKCPPLGVDSIKVEGASIKVPOKMEI 414  
RESULT 6  
Q6AYN7 PRELIMINARY; PRT; 467 AA.  
AC Q6AYN7;  
DT 25-OCT-2004 (TREMBLrel. 28, Created)  
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)  
DE Hypothetical protein (Fragment).  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
Raha S.S., Loggiano N.A., Peters G.J., Abramson R.D., Millard S.J.,  
Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,  
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,  
Krzyszinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,  
Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;



RA Director MGC Project;  
 RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC078974; AAH78974.1;  
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR00905; Peptidase\_M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
 DR InterPro; IPR002016; Peroxidase.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Osialglc\_ptide; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR Prodom; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 DR TRPOSITE; PS00435; PEROXIDASE\_1; UNKNOWN\_1.  
 DR Hypochemical protein.  
 FW NON TER 467  
 FT SEQUENCE 467 AA; 50799 MW; 474E1B1959B8AC0 CRC64;  
 SQ  
 Query Match 80.6%; Score 1116; DB 2; Length 467;  
 Best Local Similarity 88.4%; Pred. No. 3.8e-87;  
 Matches 214; Conservative 10; Mismatches 18; Indels 0; Gaps 0;

QY 1 MEAHLITRLTNKVEFPFLVLLISGHCCLALVGVSPFLIGKSLDIAPGMDLVARR 60  
 DB 148 MEAHLITRLTNKVEFPFLVLLISGHCCLALVGVSPFLIGKSLDIAPGMDLVARR 207  
 QY 61 LSLIKPECSGSGKAIIEHLAKQGNRFFDIKPLHAKNCDPFTGLQHTDKIMKK 120  
 DB 208 LSLIKPECSGSGKAIIEHLAKQGNRFFDIKPLHAKNCDPFTGLQHTDKIMKK 267  
 QY 121 EKEGIEKGQILSSADIAATVQHTMACLVKTRTHRALIFCKORDLPONNAVLVASGV 180  
 DB 268 EKEGIEKGQILSSADIAATVQHTMACLVKTRTHRALIFCKORDLPONNAVLVASGV 327  
 QY 181 ASNFYIRALBELTNATQCTLCPPRLCTDNGIMIAMNGIERLRAGGILHDEIGRIYE 240  
 DB 328 ASNFYIRALBELTNATQCTLCPPRLCTDNGIMIAMNGIERLRAGGILHDEIGRIYE 387  
 QY 241 PK 242  
 DB 388 PK 389

RESULT 7  
 O96NH5 PRELIMINARY; PRT; 364 AA.  
 AC O96NH5;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Hypothetical protein FLJ30879.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX PubMed=14702039; DOI=10.1038/ng1285;  
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 RA Makatsuru A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
 RA Sekine M., Ohtsushi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,  
 RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
 RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
 RA Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M., Yamazaki M.,  
 RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,  
 RA Tanai H., Kimachi N., Watanabe M., Hirotsuka S., Chiba Y., Ishida S.,  
 RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Houta T., Kusano J.,  
 RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,  
 RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,

RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
 RA Yoshikawa Y., Matsunaga H., Ichihara T., Shiohara N., Sano S.,  
 RA Moriya S., Komiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
 RA Hisigaki H., Watanabe T., Sugiyama A., Takemoto M., Kakawami B.,  
 RA Fujimori Y., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
 RA Fujimori Y., Komiyama M., Tashiro H., Taniguchi A., Fujimori Y.,  
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y.,  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Ohtsuka R., Wakakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togeishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 RT cDNAs";  
 RL Nat. Genet. 36:40-45(2004).  
 DR EMBL; AK055441; BAB70923.1; -.  
 DR MEROPS; M22.004; -.  
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008270; F:peptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR00905; Peptidase\_M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Osialglc\_ptide; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR Prodom; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 DR TrEMBLrel.  
 KW Peptidase.  
 SQ SEQUENCE 364 AA; 39528 MW; E0B605A07D0EC3D6 CRC64;  
 Query Match 79.8%; Score 1105; DB 2; Length 364;  
 Best Local Similarity 98.6%; Pred. No. 2.5e-86;  
 Matches 214; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MEAHLITRLTNKVEFPFLVLLISGHCCLALVGVSPFLIGKSLDIAPGMDLVARR 60  
 DB 148 MEAHLITRLTNKVEFPFLVLLISGHCCLALVGVSPFLIGKSLDIAPGMDLVARR 207  
 QY 61 LSLIKPECSGSGKAIIEHLAKQGNRFFDIKPLHAKNCDPFTGLQHTDKIMKK 120  
 DB 208 LSLIKPECSGSGKAIIEHLAKQGNRFFDIKPLHAKNCDPFTGLQHTDKIMKK 267  
 QY 121 EKEGIEKGQILSSADIAATVQHTMACLVKTRTHRALIFCKORDLPONNAVLVASGV 180  
 DB 268 EKEGIEKGQILSSADIAATVQHTMACLVKTRTHRALIFCKORDLPONNAVLVASGV 327  
 QY 181 ASNFYIRALBELTNATQCTLCPPRLCTDNGIMIA 217  
 DB 328 ASNFYIRALBELTNATQCTLCPPRLCTDNGIMIA 364

RESULT 8  
 O8JFW3 PRELIMINARY; PRT; 404 AA.  
 AC O8JFW3;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE S1d221013.4 (Novel glycoprotease).  
 GN Name=d722B1.4.6;  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Brachydanio; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 NC NCB1\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Babbage A.;  
 RA Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL591593; CAD43471.1; -.  
DR MEROPS; M22.004; -.  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR009180; Peptidase M22.  
DR InterPro; IPR009180; Peptidase M22.  
DR Pfam; PF00814; Peptidase M22; 1.  
DR PIRSF; PIRSF004537; O-sialic\_acid; 1.  
DR PRINTS; PR00789; O-sialic\_acid; 1.  
DR ProDom; PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs; TIGR00329; gcp; 1.  
KM PROTEASE.  
SQ SEQUENCE 404 AA; 43956 MW; 3A6BDD050737E35 CRC64;  
Query Match 62.0%; Score 859; DB 2; Length 404;  
Best Local Similarity 62.3%; Pred. No. 3,4e-65;  
Matches 165; Conservative 38; Mismatches 62; Indels 0; Gaps 0;  
QY 1 MEAHLATRLTNKVEPFLVLLISGGHCLALVGVSDFLGKSLDIAFGMDLVKVR 60  
DB 136 MEAHLATRLTNKVEPFLVLLISGGHCLALVGVSDFLGKSLDIAFGMDLVKVR 195  
QY 61 LSLIKHEPCSTMSGKATIEHLAKOGNRFHPDIKPELHAKNCDSPFTGLQHTVDKIIMKK 120  
DB 196 LSLIKHEPCSTMSGKATIEHLAKOGNRFHPDIKPELHAKNCDSPFTGLQHTVDKIIMKK 255  
QY 121 EKEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCCKORDLPONNAVIVASGV 180  
DB 256 EKEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCCKORDLPONNAVIVASGV 315  
QY 181 ASNFYIRRAELITNAOTCTLLCPPLRCTDNGMIAMNGIERLRAGLIDIGIRYE 240  
DB 316 ASNFYIRRAELITNAOTCTLLCPPLRCTDNGMIAMNGIERLRAGLIDIGIRYE 375  
QY 241 PKCPGVDSISKEVGEASIKVPOLKM 265  
DB 376 PKCPGVDSISKEVGEASIKVPOLKM 400

RESULT 9  
Q8JFR7 PRELIMINARY; PRT; 404 AA.  
ID Q8JFR7  
AC Q8JFR7  
DT 01-OCT-2002 (Tremblrel. 22, Created)  
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DE 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE SI:d72B14.6 (Novel glycoprotease).  
GN Name=SI:d72B14.6;  
OS Brachydanio rerio (zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Clark G.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL672217; CAD43443.1; -.  
DR MEROPS; M22.004; -.  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR009180; Peptidase M22.  
DR InterPro; IPR009180; Peptidase M22.  
DR Pfam; PF00814; Peptidase M22; 1.  
DR PIRSF; PIRSF004537; O-sialic\_acid; 1.  
DR PRINTS; PR00789; O-sialic\_acid; 1.  
DR ProDom; PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs; TIGR00329; gcp; 1.  
KM PROTEASE.

SQ SEQUENCE 404 AA; 44027 MW; 6FE98653A651860F CRC64;  
Query Match 61.7%; Score 855; DB 2; Length 404;  
Best Local Similarity 61.9%; Pred. No. 7,6e-65;  
Matches 164; Conservative 39; Mismatches 62; Indels 0; Gaps 0;  
QY 1 MEAHLATRLTNKVEPFLVLLISGGHCLALVGVSDFLGKSLDIAFGMDLVKVR 60  
DB 136 MEAHLATRLTNKVEPFLVLLISGGHCLALVGVSDFLGKSLDIAFGMDLVKVR 195  
QY 61 LSLIKHEPCSTMSGKATIEHLAKOGNRFHPDIKPELHAKNCDSPFTGLQHTVDKIIMKK 120  
DB 196 LSLIKHEPCSTMSGKATIEHLAKOGNRFHPDIKPELHAKNCDSPFTGLQHTVDKIIMKK 255  
QY 121 EKEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCCKORDLPONNAVIVASGV 180  
DB 256 EKEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCCKORDLPONNAVIVASGV 315  
QY 181 ASNFYIRRAELITNAOTCTLLCPPLRCTDNGMIAMNGIERLRAGLIDIGIRYE 240  
DB 316 ASNFYIRRAELITNAOTCTLLCPPLRCTDNGMIAMNGIERLRAGLIDIGIRYE 375  
QY 241 PKCPGVDSISKEVGEASIKVPOLKM 265  
DB 376 PKCPGVDSISKEVGEASIKVPOLKM 400

RESULT 10  
Q7Q918 PRELIMINARY; PRT; 401 AA.  
ID Q7Q918  
AC Q7Q918  
DT 01-MAR-2004 (Tremblrel. 26, Created)  
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)  
DE 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE AGCP14990 (Fragment).  
GN Name=AGCP14990; ORFNames=ENSAANG0000007922;  
OS Anopheles gambiae str. BEST.  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
OX NCBI\_TaxID=180454;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=BEST;  
RC Anopheles Genome Sequencing Consortium;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AAA801008900; BAA09387.1; -.  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR009180; Peptidase M22.  
DR InterPro; IPR009180; Peptidase M22.  
DR Pfam; PF00814; Peptidase M22; 1.  
DR PIRSF; PIRSF004537; O-sialic\_acid; 1.  
DR PRINTS; PR00789; O-sialic\_acid; 1.  
DR ProDom; PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs; TIGR00329; gcp; 1.  
FT NON TER 1  
SQ SEQUENCE 401 AA; 43849 MW; CC9426723D3FD4F1 CRC64;  
Query Match 32.7%; Score 453.5; DB 2; Length 401;  
Best Local Similarity 38.4%; Pred. No. 1,9e-30;  
Matches 106; Conservative 50; Mismatches 109; Indels 11; Gaps 4;  
QY 1 MEAHLATRLTNKVEPFLVLLISGGHCLALVGVSDFLGKSLDIAFGMDLVKVR 60  
DB 119 MEAHLATRLTNKVEPFLVLLISGGHCLALVGVSDFLGKSLDIAFGMDLVKVR 178  
QY 61 LSLIKHEPCSTMSGKATIEHLAKOGNRFHPDIKPELHAKNCDSPFTGLQHTVDKIIMKK 115  
DB 179 LSLIKHEPCSTMSGKATIEHLAKOGNRFHPDIKPELHAKNCDSPFTGLQHTVDKIIMKK 236

```
QY 116 TIMKEBEBGKQOISSADIAATVQHTACHLVKTRTHAILFCQORDL--PQNNAV 173
DB 237 HIERESTLHAPDALPDYEAFCACFLKGVTRHMLHRTORALEYCEBRKFLPSDAEPHRS 296
QY 174 LVASGVASNYIRALIELTNATQCTLLCPRPRLCTNGIMIANGIERTLRA--GIGLIL 231
DB 297 LVASGVACNDVIRNALSMAAQGYSTYRPPKCLCTNGTWTIANNGEKLAKDTAEMT 356
QY 232 HDIEGIRYEPKCPGLVDISKVEGASIKYPOLKMEI 267
DB 357 TKYEQVDISGKCPICGDSLIDVTKANLACKAKVADI 392

RESULT 11
Q96086 PRELIMINARY; PRT; 323 AA.
AC 096086;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-MAR-2004 (TReMBLrel. 19, Last sequence update)
DE 1D37221P.
GN ORFNames=CG14231;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota, Metazoa, Arthropoda, Hexapoda, Insecta, Pterygota,
OC Neoptera, Endopterygota, Diptera, Brachycera, Muscomorpha;
OC Ephydroidea, Drosophilidae, Drosophila.
NCBI_TaxID=7227;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacled J., Paragas V., Park S., Prounenavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Ceiniker S.;
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY051882; AK93306.1; -
DR FlyBase; FBgn001060; CG14231.
DR GO; GO:0008450; F.O-sialoglycoprotein endopeptidase activity; IEA.
DR GO; GO:0008270; F.zinc ion binding; IEA.
DR GO; GO:0006508; P.proteinolysis and peptidolysis; IEA.
DR InterPro; IPR009095; Peptidase M22.
DR InterPro; IPR009180; Peptidase M22.
DR Pfam; PF00814; Peptidase M22; 1.
DR PIRSF; PIRSF004537; Osa1g1c_prcds; 1.
DR PRINTS; PR00789; OSIALOPTASE.
SQ SEQUENCE 323 AA; 35828 MW; BBDDB54D9A2BF35A CRC64;

Query Match 30.0%; Score 415.5; DB 2; Length 323;
Best Local Similarity 38.2%; Pred. No. 2.7e-27;
Matches 100; Conservative 43; Mismatches 104; Indels 15; Gaps 6;

QY 1 MEAHALTRILTN--KVEFPLVLLISGGCHLALVQGYSDFLLEKSIDIAPGMDLVYA 58
DB 50 MEAHALQARHEHPQIGVPLCLLASGHCQLVANGRGRLTLTGOTLDDAPGEAFDXIG 109
QY 59 RLISLTHPEPCSTMSGGKALIEHLAK--QGNRFHFDIKPPLHAKNCDSFPTGLQHTDKII 117
DB 110 RLRLHLTLPEYRLNNGRAIEHAQLASDPLAYEPFLPLAQORNCNFSFAGIKNNSPFAI 169
QY 118 MKKEKEGIEKQGLSSAADIAATVQHTACHLVKTRTHAILFC--KQRDLIPQNNATLV 175
DB 170 RABBRARTRPDGIVSYGDCAGULSSVSHRLHRTORALEYCLPRLQFGDTPPTLV 229
QY 176 ASGGVANSFYIRALIELTNATQCTLLCPRPRLCTDNGIMIANGIERTLRA--GIGLIL 235
DB 230 MSGGVANNDIAIYANIEHLAAQYGRSRFRPSKRYCSDNQVWLAMHGEVL-----LDKKE 283
QY 236 -GIRYEPKCPGLVDISKVEGASIKYPOLKMEI 266
DB 284 ASTRYD--SIDIQSAGFA 302
```

```
RESULT 12
Q9VWD6 PRELIMINARY; PRT; 409 AA.
AC 09VWD6;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE CG14231-PA.
GN ORFNames=CG14231;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota, Metazoa, Arthropoda, Hexapoda, Insecta, Pterygota;
OC Neoptera, Endopterygota, Diptera, Brachycera, Muscomorpha;
OC Ephydroidea, Drosophilidae, Drosophila.
NCBI_TaxID=7227;
RX [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazet R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Idegawa C.,
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimbel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacled J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassenaar D.A., Weinstock G.M., Weisenbach J.,
RA Williams S.M., Woodgett W., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zheng F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL "The genome sequence of Drosophila melanogaster." ;
RL Science 287:2185-2195 (2000).

[2]
RN
RP SEQUENCE FROM N.A.
RC MEDLINE=22426065; PubMed=12537568;
RA Celiker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacled J.M., Park S., Pfeiffer B.D., Richards S., Sodeglen E.J.,
RA Svirskas R., Taber P.B., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RL melanogaster euchromatic genome sequence." ;
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).

[3]
RN
RP SEQUENCE FROM N.A.
RC MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
```

RA Ashburner M., Celinker S.E.;  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 a genomics perspective";  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2242609; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik K.S.,  
 Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 Betencourt B.R., Celinker S.E., de Grey A.D., Drysdale R.A.,  
 Harris N.L., Richter C., Russo S., Schroeder A.J., Shu S.O.,  
 Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 Lewis S.E.;  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 systematic review";  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RG FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RG FlyBase;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AEO03513; AAF49008.1; -  
 DR InAct; Q9VMD6; -  
 DR FlyBase; FBgn0031060; CG14231.  
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009180; Pept M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Osialglc\_ptcds; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 SQ SEQUENCE 409 AA; 45328 MW; 9797F667D155538 CRC64;  
 Query Match 30.0%; Score 415.5; DB 2; Length 409;  
 Best Local Similarity 38.2%; Pred. No. 3.6e-27;  
 Matches 100; Conservative 43; Mismatches 104; Indels 15; Gaps 6;  
 QY 1 MEAHALTRLTN--KVERPVLVILSGHCHLAVQVSDPFLILGKSLDIAGDMLDKVA 58  
 DB 136 MEALALQARMEHPQIGYPLCLLASGGHCDLVANGPRLTLLOQTDDAGEAFDKIG 195  
 QY 59 RRLSLIKHPECSWGGKAIENLAK-QGNRFHFDIKPPLHAKNCDPSTGLOHTVDKII 117  
 DB 196 RRLRLHLPEYRLANGGRAIEHAAQLADPLAYEPFLPLAQORACNCFSPAGIKNNSFRAI 255  
 QY 118 MKKEKEGIEKQILSSAADIAATVQHTMACHLVKTRHAIIFC--KORDLDPONNAVIV 175  
 DB 256 RARERARTRPDGVIISNYGDFCAGLRVSRHLMRTQRAIEYCLLPHQRLGDPPTPLV 315  
 QY 176 ASGVASNFYRLALEITLTAOTCTLPPRLCTDNGIMAMNGIERLPAAGLGIHIE 235  
 DB 316 MSGGVANNDAIYANIEHLAAYOYGRSFRPSKRYCSDNGVIMAMHGEQI-----LQDKK 369  
 QY 236 -GIYEPKCPGLGVNISKEVGEA 256  
 DB 370 ASTRYDYD--SIDIQSAGFA 388  
 RESULT 13  
 022145 PRELIMINARY; PRT; 480 AA.  
 ID 022145  
 AC 022145; Q8VWL2;  
 DT 01-JUN-1998 (TRENBLrel. 05, Created)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)  
 DT 05-JUN-2004 (TRENBLrel. 27, Last annotation update)  
 DE Putative O-sialoglycoprotein endopeptidase (Sialoglycoprotease

DE (GCP1).  
 GN Name=At2g45270; Synonyms=GCP1;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 ON Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,  
 RA Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R.,  
 RA Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Town C.D., Kaul S.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Hauswirth K., Adamska I.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
 RA Seki M., Shim P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,  
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,  
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
 RA Seki M., Shim P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 DR EMBL; AC002338; AAB82636.2; -  
 DR EMBL; AY024338; AAK00530.1; -  
 DR EMBL; AY063864; AAL36220.1; -  
 DR EMBL; AY117283; AAM51358.1; -  
 DR PIR; E84888; E84888.  
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009095; Peptidase\_M22.  
 DR InterPro; IPR009180; Pept M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Osialglc\_ptcds; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 KW Protease.  
 SQ SEQUENCE 480 AA; 52995 MW; 20DD6A86ACCF1FPAAD CRC64;  
 Query Match 29.4%; Score 407.5; DB 2; Length 480;  
 Best Local Similarity 37.0%; Pred. No. 2.1e-26;  
 Matches 104; Conservative 37; Mismatches 93; Indels 47; Gaps 7;  
 QY 1 MEAHALTRLTN-KVERPVLVILSGHCHLAVQVSDPFLILGKSLDIAGDMLDKVAR 59  
 DB 195 MEHALVALVBEELSFPMALLISGHNILVLAHKGQYTGQYTGDDAIGAEFDTAK 254  
 QY 60 RSLIKHPECSWGGKAIENLAKQGNRFHFDIKPPLHAKNCDPSTGLOHTVDKIIK 119  
 DB 255 WGLGLDWH-----RSGPAVEALLEGDAKSVKENVPMKHKDCNFSYAGLKTQVRLAIEA 309

```
QY 120 KEKEEGIEKGQILSSA-----ADINATQHTMACHLYKTRTRAILFCCKORDLLPQNN 171
DB 310 KE-----IDAKCPVSSATNEDRRNRADIAASFQRAVALHEKCECAIDMALE---LEPSI 362
QY 172 AVIVASGVASNFYIRRALLETITNATQCTLLCPRLCTDNGIMANNGIERLRAGLGL 231
DB 363 KHWIVSGGVASNKYVRLRNINVENKMLKLVCPPLSLCTDNGVMVAMTGLEHFRVG---- 418
QY 232 HDIEGIRYE-----PKCPLGVDISKEVGEA 256
DB 419 -----RYDPPPPATPEPDYVYDRLPRMPPLGEBYAKGSEA 453
RESULT 14
Q73H71 PRELIMINARY; PRT; 335 AA.
ID Q73H71
AC Q73H71
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Peptidase, M22 family protein.
OS OrderedLocustNames=WD0699;
GN Wolbachia pipiensis wMel.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Wolbachiae; Wolbachia.
OX NCBI_Taxid=66077;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15024419;
RA Wu M., Sun L.V., Vamathevan J.J., Riegler M., Deboy R.T.,
RA Brownlie J.C., McGraw E.A., Martin W., Esser C., Ahmadijad N.,
RA Wiegand C., Madupu R., Beaman M.J., Brinkac L.M., Daugherty S.C.,
RA Durkin A.S., Kolonay J.F., Nelson W.C., Mohamoud Y., Lee P.,
RA Berry K.J., Young M.B., Utterback T.R., Weidman J.P., Nierman W.C.,
RA Paulsen I.T., Nelson K.E., Tettelin H., O'Neill S.L., Eisen J.A.,
RT "Phylogenomics of the reproductive parasite Wolbachia pipiensis wMel:
RT a streamlined genome overruled by mobile genetic elements."
RL PLOS Biol. 2:327-341(2004).
DR EMBL: AE017258; AAS14395.1; -.
DR TIGR: WD0699; -.
DR GO: GO:0008450; P:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR000905; Peptidase M22.
DR InterPro: IPR009180; Pept M22_Osialgl.
DR Pfam: PF00814; Peptidase_M22; 1.
DR PIRSF: PIRSF004537; Osialglc_ptds; 1.
DR PRINTS: PR00789; OSIALOPTASE.
DR ProDom: PD002367; Peptidase_M22; 1.
DR TIGRFAMs: TIGRFAMs0329; gcp; 1.
DR Complete proteome.
SQ SEQUENCE 335 AA; 36634 MW; F4914CDA58BC9390 CRC64;
Query Match 26.6%; Score 369; DB 2; Length 335;
Best Local Similarity 34.7%; Pred. No. 2.8e-23;
Matches 84; Conservative 46; Mismatches 86; Indels 26; Gaps 5;
QY 1 MEHAALTRLTNKEVPPFLVLLISGGHCLALVGVSDPLLGKSLDIAPGMDLDKVAR 60
DB 112 LEHALVIRLRLHEVYKFPFLVLLISGGHCOPLIADVGKYLKLGELDLDSLBAPFKVAKM 171
QY 61 LSLIKHECSTMSGKALEHLAKGNRFHPIKPLHLAKNCDFSTGLQHTVDKIINK 120
DB 172 LGL-----SYGCPPLIEKAKKGNTRFPLPRAMIKRSGCNFSFGSKTRAVKLVDEL 224
QY 121 EKEBIEKGQILSSAADIAATVQHTMACHLYKTRTRAILFCCKORDLLPQNNAVIVASGV 180
DB 225 KMSF-----QVQVQVCAFSQECISDIILDRVSNALIMBSALIKIND---FYITGCV 273
QY 191 ACNPFYIRRALLETITNATQCTLLCPRLCTDNGIMANNGIERLRAGLGL 240
DB 274 AANNFLAEKJLKHQHN---LNIFFPPNDLCTDNALIVGVTGIERLQKYN-----IDPLMFA 325
```

```
QY 241 PK 242
DB 326 PR 327
RESULT 15
Q92LH8 PRELIMINARY; PRT; 360 AA.
ID Q92LH8
AC Q92LH8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PROBABLE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57).
GN ORFNames=SMC03230;
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_Taxid=382;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;
RC STRAIN=1021;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Barut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss B., Lelaure V., Masny D.,
RA Pohl T., Portetalle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL: AL591792; CAC47657.1; -.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR000905; Peptidase_M22.
DR InterPro: IPR009180; Pept M22_Osialgl.
DR Pfam: PF00814; Peptidase_M22; 1.
DR PIRSF: PIRSF004537; Osialglc_ptds; 1.
DR PRINTS: PR00789; OSIALOPTASE.
DR ProDom: PD002367; Peptidase_M22; 1.
DR TIGRFAMs: TIGRFAMs0329; gcp; 1.
DR Complete proteome; Hydrolase.
SQ SEQUENCE 360 AA; 37906 MW; A07F946AB562EAB6 CRC64;
Query Match 25.7%; Score 356.5; DB 2; Length 360;
Best Local Similarity 35.6%; Pred. No. 3.5e-22;
Matches 95; Conservative 38; Mismatches 91; Indels 43; Gaps 7;
QY 1 MEHAALTRLTNKEVPPFLVLLISGGHCLALVGVSDPLLGKSLDIAPGMDLDKVAR 60
DB 116 LEHALVIRLRLHEVYKFPFLVLLISGGHCOPLIADVGKYLKLGELDLDSLBAPFKVAKM 175
QY 61 LSLIKHECSTMSGKALEHLAKGNRFHPIKPLHLAKNCDFSTGLQHTVDKIINK 120
DB 176 LGL-FYP-----GGPAYVRAAQAQNAERFDPRLVGPARDIFSFGKTRVQAQSL 228
QY 111 -HTVDKIINKKEBIEKGQILSSAADIAATVQHTMACHLYKTRTRAILFCCKORDLLQ 169
DB 229 GPTVDQDI-----ADVCAFSQRAISRITLDRVGRGKRFR-ADFAV 269
QY 170 NNAVIVASGVASNFYIRRALLETITNATQCTLLCPRLCTDNGIMANNGIERLRAGL 229
DB 270 DQPALVAVGVANQTLRRTLOSICDHEGFRFAPLQVCTDNAAAMIAAGABRLAAGL- 328
QY 230 ILHDIEGIRYEBKC--PLGVDISKEVG 254
DB 329 ---PADGIDAAPRSRMPDSEAKALIG 352
Search completed: November 10, 2005, 16:52:02
Job time : 65.7185 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:51:52 ; Search time 21.4761 Seconds  
(without alignments)  
928.069 Million cell updates/sec

```

Title:      US-10-649-273-2_COPY_148_414
Perfect score: 1385
Sequence:   1 MEAHALTIRLTNKKVEPPPLV.....DISKEVGASIKVPQLKMEI 267

```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match	0%
Minimum Match	100%	

```
Database : Issued Patents AA:*
1 : /cgn2_6/pdataa/1/aa/5A_COMB pep:*
2 : /cgn2_6/pdataa/1/aa/5F_COMB pep:*
3 : /cgn2_6/pdataa/1/aa/6A_COMB pep:*
4 : /cgn2_6/pdataa/1/aa/6B_COMB pep:*
5 : /cgn2_6/pdataa/1/aa/PRTUS_COMB pep:*
6 : /cgn2_6/pdataa/1/aa/backfilea1 pep:
```

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1385	100.0	267	4	US-10-067-443-22	Sequence 22, Appl
2	1385	100.0	414	4	US-10-067-443-2	Sequence 2, Appl
3	1358.5	98.1	439	4	US-10-067-443-19	Sequence 19, Appl
4	409.5	29.6	463	4	US-10-067-443-3	Sequence 3, Appl
5	307	22.2	350	4	US-09-540-236-2726	Sequence 1726, Ap
6	281.5	21.0	401	4	US-09-252-991A-17372	Sequence 17372, A
7	280.5	20.3	357	4	US-09-543-681A-6513	Sequence 6513, Ap
8	279	20.1	342	1	US-08-087-797-3	Sequence 3, Appl
9	276	19.9	421	4	US-10-067-443-4	Sequence 4, Appl
10	276	19.5	421	4	US-10-067-443-28	Sequence 28, Appl
11	270.5	19.5	325	1	US-08-087-797-2	Sequence 2, Appl
12	268.5	19.4	343	4	US-09-489-039A-9221	Sequence 9221, Ap
13	261	18.8	363	4	US-09-107-532A-6609	Sequence 6609, Ap
14	254	18.3	336	3	US-08-987-121A-4	Sequence 4, Appl
15	250	18.1	335	3	US-08-961-083-52	Sequence 52, Appl
16	250	18.1	335	4	US-09-536-784-52	Sequence 52, Appl
17	248	17.9	336	3	US-09-066-512-2	Sequence 2, Appl
18	246	17.8	336	4	US-09-583-110-4857	Sequence 4857, Ap
19	243	17.5	327	4	US-10-067-443-5	Sequence 5, Appl
20	241	17.4	336	4	US-09-107-433-4221	Sequence 4221, Ap
21	236	17.0	273	4	US-09-170-279-728	Sequence 728, Appl
22	236	17.0	366	4	US-09-314-000C-4956	Sequence 4956, Ap
23	223	16.8	368	3	US-09-134-001C-3909	Sequence 3909, Ap
24	223	15.5	341	3	US-09-189-624-2	Sequence 2, Appl
25	218	15.7	344	4	US-09-198-452A-213	Sequence 213, Appl
26	218	15.7	360	4	US-09-438-185A-196	Sequence 196, Appl
27	211.5	15.3	344	4	US-09-602-777A-148	Sequence 148, Appl

28	194	14.0	340	4	US-09-0267-443-6	Sequence 6, Appl
29	181.5	13.1	133	4	US-09-3282-355-4387	Sequence 4387, Appl
30	169	12.2	292	4	US-09-7242-623-81	Sequence 81, Appl
31	157	11.3	214	4	US-09-3282-352-6609	Sequence 4609, Appl
32	94.5	6.8	480	4	US-09-583-110-5050	Sequence 5050, Appl
33	94.5	6.8	481	4	US-09-107-423-3197	Sequence 3197, Appl
34	83	6.0	42	4	US-09-902-540-13841	Sequence 13841, Appl
35	81.5	5.9	1087	1	US-08-264-002-5	Sequence 5, Appl
36	81	5.8	1072	4	US-09-902-540-15572	Sequence 15572, Appl
37	79	5.7	328	4	US-09-710-279-9308	Sequence 3008, Appl
38	79	5.7	336	3	US-09-134-001C-5549	Sequence 5549, Appl
39	79	5.7	445	3	US-08-083-945C-2	Sequence 2, Appl
40	79	5.7	445	3	US-08-083-945C-7	Sequence 7, Appl
41	79	5.7	578	4	US-09-252-991A-11318	Sequence 11318, Appl
42	79	5.7	1137	4	US-09-538-092-968	Sequence 968, Appl
43	78.5	5.7	503	4	US-09-252-991A-22790	Sequence 22790, Appl
44	78.5	5.7	1658	4	US-08-603-049A-13	Sequence 13, Appl
45	78.5	5.7	1658	3	US-09-170-996-13	Sequence 13, Appl

## ALIGNMENTS

RESULT 1  
US-10-067-443-22  
; Sequence 22, Application US/10067443

; GENERAL INFORMATION:

TITLE OF INVENTION: POLYNUCLEOTIDE ENCC  
TITLE OF INVENTION: SPINNY COCD MD 1

FILE REFERENCE: D0073 NE  
CITIZEN APPLICATION NUMB

CURRENT FILING DATE: 2003-03-05  
CURRENT APPLICATION NUMBER: US/10/067,443

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

```

; NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.0

```

; SOFTWARE: F  
; SEO ID NO 22

; LENGTH: 267

TYPE: PRT

ORGANISM: homo sapiens

US-10-067-443-22

Query Match	100.0%;	Score 1385;	DB 4;	Length 267;
-------------	---------	-------------	-------	-------------

	Matches	267	Conservative	0	Mismatches	0	Indels	0	Gaps	0
Qy	1	MEAAATLIRLTNKNKPEPPLVLLISGGHCLLALVQGVSPFLILGKSLDLIARGMULDPVAR	60							
Db	1	MEAAATLIRLTNKNKPEPPLVLLISGGHCLLALVQGVSPDLILGKSLDLIARGMULDPVAR	60							
Qy	61	LSLIRKHECSTMSGAKAIEHLAKQGNRHPDIKPLIHAANKDFSFTGLQHTYTDKIIMKX	120							
Db	61	LSLIRKHECSTMSGAKAIEHLAKQGNRHPDIKPLIHAANKDFSFTGLQHTYTDKIIMKX	120							
Qy	121	EKEGIEKGGQILSSAADIATATVOHTMACHLVKRTTRAILFCQKQDILLPONNAVLAASGV	180							
Db	121	EKEGIEKGGQILSSAADIATATVOHTMACHLVKRTTRAILFCQKQDILLPONNAVLAASGV	180							
Qy	181	ASNFYIRPALEILTNAQTCLLCPBPRLCTDNGIMIANNGIERLPAAGLILHIDIEGIRYE	240							
Db	181	ASNFYIRPALEILTNAQTCLLCPBPRLCTDNGIMIANNGIERLPAAGLILHIDIEGIRYE	240							
Qy	241	PKCPPLGVDISKVEGEASIKVPQLKMEI	267							
Db	241	PKCPPLGVDISKVEGEASIKVPQLKMEI	267							



```
; Sequence 2, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-067-443-2
```

```
Query Match          100.0%; Score 1385; DB 4; Length 414;
Best Local Similarity 100.0%; Pred. No. 7.3e-160;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 MEAHALTIRLTNKKVEPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 60
Db 148 MEAHALTIRLTNKKVEPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 207
Qy 61 LSLIKHPECSTWGGKAEIHLAKQGNRPFHDIKPPLHAKNCDPFTGLQHTVDKXIMKK 120
Db 208 LSLIKHPECSTWGGKAEIHLAKQGNRPFHDIKPPLHAKNCDPFTGLQHTVDKXIMKK 267
Qy 121 EKEGIEKQGLSSAADIAATVQHTMACHLVKTRATILFCQKQRLPQNNNAVVASGV 180
Db 268 EKEGIEKQGLSSAADIAATVQHTMACHLVKTRATILFCQKQRLPQNNNAVVASGV 327
Qy 181 ASNFYIRRALIILTNATQCTLLCPPRLCTDNGMIANNGIERLRAGGILHDIGIRY 240
Db 328 ASNFYIRRALIILTNATQCTLLCPPRLCTDNGMIANNGIERLRAGGILHDIGIRY 387
Qy 241 PKCPLGVDISKVEGEASIKVPOLKMEI 267
Db 388 PKCPLGVDISKVEGEASIKVPOLKMEI 414
```

```
RESULT 3
US-10-067-443-19
; Sequence 19, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-443-19
```

```
Query Match          98.1%; Score 1358.5; DB 4; Length 439;
Best Local Similarity 91.1%; Pred. No. 1.4e-156;
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;
```

```
Qy 1 MEAHALTIRLTNKKVEPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 60
Db 148 MEAHALTIRLTNKKVEPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 207
Qy 61 LSLIKHPECSTWGGKAEIHLAKQGNRPFHDIKPPLHAKNCDPFTGLQHTVDKXIMKK 120
Db 208 LSLIKHPECSTWGGKAEIHLAKQGNRPFHDIKPPLHAKNCDPFTGLQHTVDKXIMKK 267
Qy 121 EKEGIEKQGLSSAADIAATVQHTMACHLVKTRATILFCQKQRLPQNNNAVVASGV 180
Db 268 EKEGIEKQGLSSAADIAATVQHTMACHLVKTRATILFCQKQRLPQNNNAVVASGV 327
Qy 181 ASNFYIRRALIILTNATQCTLLCPPRLCTDNGMIANNGIERLRAGGILHDIGIRY 240
Db 328 ASNFYIRRALIILTNATQCTLLCPPRLCTDNGMIANNGIERLRAGGILHDIGIRY 387
Qy 241 PKCPLGVDISKVEGEASIKVPOLKMEI 267
Db 388 IAMNGIERLRAGGILHDIGIRYEPKCPPLGVDISKVEGEASIKVPOLKMEI 439
```

```
Query Match          100.0%; Score 1385; DB 4; Length 414;
Best Local Similarity 100.0%; Pred. No. 7.3e-160;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 MEAHALTIRLTNKKVEPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 60
Db 148 MEAHALTIRLTNKKVEPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 207
Qy 61 LSLIKHPECSTWGGKAEIHLAKQGNRPFHDIKPPLHAKNCDPFTGLQHTVDKXIMKK 120
Db 208 LSLIKHPECSTWGGKAEIHLAKQGNRPFHDIKPPLHAKNCDPFTGLQHTVDKXIMKK 267
Qy 121 EKEGIEKQGLSSAADIAATVQHTMACHLVKTRATILFCQKQRLPQNNNAVVASGV 180
Db 268 EKEGIEKQGLSSAADIAATVQHTMACHLVKTRATILFCQKQRLPQNNNAVVASGV 327
Qy 181 ASNFYIRRALIILTNATQCTLLCPPRLCTDNGMIANNGIERLRAGGILHDIGIRY 240
Db 328 ASNFYIRRALIILTNATQCTLLCPPRLCTDNGMIANNGIERLRAGGILHDIGIRY 387
Qy 241 PKCPLGVDISKVEGEASIKVPOLKMEI 267
Db 388 IAMNGIERLRAGGILHDIGIRYEPKCPPLGVDISKVEGEASIKVPOLKMEI 439
```

```
RESULT 5
US-09-540-236-2726
; Sequence 2726, Application US/09540236
; Patent No. 6673910
```

```
Query Match          98.1%; Score 1358.5; DB 4; Length 439;
Best Local Similarity 91.1%; Pred. No. 1.4e-156;
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;
```



GENERAL INFORMATION:  
APPLICANT: Gary H. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR  
FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS  
CURRENT APPLICATION NUMBER: 2709.2005-001  
CURRENT FILING DATE: US/09/540,236  
NUMBER OF SEQ ID NOS: 3840  
SEQ ID NO 2726  
LENGTH: 350  
TYPE: PRT  
ORGANISM: M.catarhalis  
US-09-540-236-2726

Query Match 22.2%; Score 307; DB 4; Length 350;  
Best Local Similarity 33.3%; Pred. No. 1.7e-28;  
Matches 84; Conservative 37; Mismatches 91; Indels 40; Gaps 7;

QY 1 MEAHATIRLT-NKVEPFLVLLISGHCILALVQVSDFLLGKSLDIAPGMDLVKVA 58  
DB 119 MEGHILAPLSDPSPFPVCLVSGHTLVADGAVQIIGESIDDAVGECDKTA 178  
QY 59 RLSLIKHEPCSTMSGKAIIEHLAKOGRFHFDIKRPLHAKKNDSPFTGLOHTDKIM 118  
DB 179 KMLKL-PYP-----GGPNIEKLAKNPRAVELPRPQHR-KGLDFSPSGKTAIHNLIK 230  
QY 119 KEKEGIEKQILSSAADIAATVQHTMACHLVKTRTHALIFCKORDLPQNNAVLVASG 178  
DB 231 DTPNQSDP-----ARRADIASPEYAVVDLVKCKTKALQMTIRO-----LVVAG 277  
QY 179 GVASNFYIRALIELTNATQCTLLCPPRCTDNGIMIANNGIERLRAG-----227  
DB 278 GVSANQTLRRLTETRLQIDASVYAPTELCTDNGAMIAVAGFRLSRGQSDDLAVRCIP 337  
QY 228 -----LGILHD 233  
DB 338 RWDMTMLGIEYD 349

RESULT 6  
US-09-252-991A-17372  
Sequence 17372, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 17372  
LENGTH: 401  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17372

Query Match 21.0%; Score 291.5; DB 4; Length 401;  
Best Local Similarity 35.4%; Pred. No. 1.6e-26;  
Matches 87; Conservative 37; Mismatches 99; Indels 23; Gaps 9;

QY 1 MEAHATIRLT-NKVEPFLVLLISGHCILALVQVSDFLLGKSLDIAPGMDLVKVA 59  
DB 172 MEGHILAPLSDPSPFPVCLVSGHTLVADGAVQIIGESIDDAVGECDKTA 231  
QY 60 RLSLIKHEPCSTMSGKAIIEHLAKOGRFHFDIKRPLHAKKNDSPFTGLOHTDKIM 119  
DB 232 LIGL-GYP-----GGPEIARLAEGRTPGRFVPRPMTDRPGDPSFGKTKPTLN-TWQ 283

QY 120 KEKEGIEKQILSSAADIAATVQHTMACHLVKTRTHALIFCKORDLPQNNAVLVASG 179  
DB 284 RCVEAGDDSEQ---TRCDIALAFQTAVETLLIKRRAL---KOTGL--KN---LVYAGG 332  
QY 180 VASNFYIRALIELTNATQCTLLCPPRCTDNGIMIANNGIERLRAGLGILHDIGIRY 239  
DB 333 VASQALNSGLEKMLGEMKQGVYVAPRPTDNGAMIAVAGCQRLIAG---QHDGPAISV 389  
QY 240 EEPKCP 245  
DB 390 QPRWPM 395

RESULT 7  
US-09-543-681A-6513  
Sequence 6513, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
FILE REFERENCE: 2709.1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 6513  
LENGTH: 357  
TYPE: PRT  
ORGANISM: Proteus mirabilis  
US-09-543-681A-6513

Query Match 20.3%; Score 280.5; DB 4; Length 357;  
Best Local Similarity 32.7%; Pred. No. 3e-25;  
Matches 81; Conservative 35; Mismatches 101; Indels 31; Gaps 8;

QY 1 MEAHATIRLT-NKVEPFLVLLISGHCILALVQVSDFLLGKSLDIAPGMDLVKVA 59  
DB 129 MEGHILAPLSDPSPFPVCLVSGHTLVADGAVQIIGESIDDAVGECDKTA 188  
QY 60 RLSLIKHEPCSTMSGKAIIEHLAKOGRFHFDIKRPLHAKKNDSPFTGLOHTDKIM 119  
DB 189 LIGL-DYP-----GGPVLSSKWAQGVGGRFVPRPMTDRPGDPSFGKTKPTLN 241  
QY 120 KEKEGIEKQILSSAADIAATVQHTMACHLVKTRTHALIFCKORDLPQNNAVLVASG 178  
DB 242 NDDSE-----QTRADIAFAFEDAVVDTLAIKGRRA-----LEQTFKRLVMAG 284  
QY 179 GVASNFYIRALIELTNATQCTLLCPPRCTDNGIMIANNGIERLRAG-LGILHDIGI 237  
DB 285 GVSANQTLRRLTETRLQIDASVYAPTELCTDNGAMIAVAGFRLSRGQSDDLAVRCIP 340  
QY 238 RYBPKCP 245  
DB 341 TVRPRWPL 348

RESULT 8  
US-08-087-797-3  
Sequence 3, Application US/08087797  
Patent No. 5543312  
GENERAL INFORMATION:  
APPLICANT: Mellors, Alan  
APPLICANT: Lo, Reggie Y.C.  
APPLICANT: Abdullah, Khalid M.  
TITLE OF INVENTION: Pasteurella Haemolytica  
TITLE OF INVENTION: Glycoprotease  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: Bell, Seltzer, Park & Gibson, P.A.  
STREET: 1211 Bast Morehead Street,

```

; CITY: Charlotte
; STATE: No. 5543312th Carolina
; COUNTRY: United States
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/087,797
; FILING DATE: 14-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Layton, Jr., Samuel G.
; REGISTRATION NUMBER: 22807
; REFERENCE/DOCKET NUMBER: 3374-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 704 377 1561
; TELEFAX: 704 334 2014
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-087-797-3

```

Query Match 20.1%; Score 279; DB 1; Length 342;

Best Local Similarity 33.6%; Pred. No. 4.2e-25;

Matches 83; Conservative 32; Mismatches 108; Indels 24; Gaps 8;

```

QY 1 MEAAHTIRLT-TNKVEPPELVLLISGGHCLLALVGVSDPFLILGKSLDIAPGMDLKVAR 59
DB 112 MEGHLLPMDNDPEPFFVALVLSGGHTQLISVGIQVELLGSIDPAAGEAPDKTK 171
QY 60 RLSTIKPEPCSTMSGKKAIEHLAKQGN--RHFIDIKPPLHAKNCDPSTFGLOHYTDKI 119
DB 172 LGLG-DYF-----GGPLSTGMAAGTGRVFPFPMTRDRLDPSFGSLKTFPAANTIRD 224
QY 120 KEKEEGIEKGQILSSAADIAATVQHTMACHLVKTRTRAILFCQORDLLPQNN-AVLVAVSG 178
DB 225 NXXXXXGTDDQ---TRADIAFAFEDAVVDITMIKCKRA-----LDQTEFKRLVWAG 272
QY 179 GVAENFYIRRALBITLTAATCTLLCPPRRLCTDNGIMIANNGIERLRAGLILHDIEGR 238
DB 273 GVSANRTLRKALAEWMKKRGGEVYFARBEFTDNGAMTAYAGVAFKA--GATADL-GVS 329
QY 239 YEPKCP 245
DB 330 VRPRWPL 336

```

RESULT 9

US-10-067-443-4

; Sequence 4, Application US/10067443

; Patent No. 6642041

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I

; TITLE OF INVENTION: SPINAL CORD, MP-1

; FILE REFERENCE: D0073 NP

; CURRENT APPLICATION NUMBER: US/10/067,443

; PRIOR FILING DATE: 2002-02-05

; PRIOR APPLICATION NUMBER: US 60/266,518

; PRIOR FILING DATE: 2001-02-05

; PRIOR APPLICATION NUMBER: US 60/282,814

; PRIOR FILING DATE: 2001-04-10

; NUMBER OF SEQ ID NOS: 71

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 4

; LENGTH: 421

```

; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; US-10-067-443-4

```

Query Match 19.9%; Score 276; DB 4; Length 421;

Best Local Similarity 30.4%; Pred. No. 1.4e-24;

Matches 85; Conservative 52; Mismatches 107; Indels 36; Gaps 8;

```

QY 1 MEAAHTIRLT-TNKVEPPELVLLISGGHCLLALVGVSDPFLILGKSLDIAPGMDLKVAR 59
DB 134 MRAHALSTILVDDSVRPPSAVLISGGHALISVADEVKFKLYGVGVSSPSCIDKVAR 193
QY 60 RLSTIKPEPCSTMSGKKAIEHLAKQGN--RHFIDIKPPLHAKNCDPSTFGLOHYTDKI 116
DB 194 QLGDL-GSEFDDGIHGAAVEILASRASADGHLRYPIFLPNVRKANNFPQIKGSYNTLLE 252
QY 117 IMKKEEGIEKGQILSSAADIAATVQHTMACHLVKTRTRAILFCQORDLLPQNNAVLVA 176
DB 253 RLKNSETSID-----IPDFCASLQNTVAHHSKLIHFESLSQEKLPQO---LVI 302
QY 177 SGVANSFYIRRALBITLTAATCTLLCPPRRLCTDNGIMIANNGIERLRAGLILHDIEG 236
DB 303 GGVANAOYIIFGAIKSLAAHNVTTIKVLISCTDNAMEIAYSGL-----LMLVNRSEA 356
QY 237 IRYEP-----KCPLGVDISKEVGEASIKVPQLKM 265
DB 357 IWRPNDIPDTIYAHARSDIGTDASSEI---IDTPRRKL 392

```

RESULT 10

US-10-067-443-28

; Sequence 28, Application US/10067443

; Patent No. 6642041

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I

; TITLE OF INVENTION: SPINAL CORD, MP-1

; FILE REFERENCE: D0073 NP

; CURRENT APPLICATION NUMBER: US/10/067,443

; PRIOR FILING DATE: 2002-02-05

; PRIOR APPLICATION NUMBER: US 60/266,518

; PRIOR FILING DATE: 2001-02-05

; PRIOR APPLICATION NUMBER: US 60/282,814

; NUMBER OF SEQ ID NOS: 71

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 28

; LENGTH: 421

; TYPE: PRT

; ORGANISM: Caenorhabditis elegans

US-10-067-443-28

Query Match 19.9%; Score 276; DB 4; Length 421;

Best Local Similarity 30.4%; Pred. No. 1.4e-24;

Matches 85; Conservative 52; Mismatches 107; Indels 36; Gaps 8;

```

QY 1 MEAAHTIRLT-TNKVEPPELVLLISGGHCLLALVGVSDPFLILGKSLDIAPGMDLKVAR 59
DB 134 MRAHALSTILVDDSVRPPSAVLISGGHALISVADEVKFKLYGVGVSSPSCIDKVAR 193
QY 60 RLSTIKPEPCSTMSGKKAIEHLAKQGN--RHFIDIKPPLHAKNCDPSTFGLOHYTDKI 116
DB 194 QLGDL-GSEFDDGIHGAAVEILASRASADGHLRYPIFLPNVRKANNFPQIKGSYNTLLE 252
QY 117 IMKKEEGIEKGQILSSAADIAATVQHTMACHLVKTRTRAILFCQORDLLPQNNAVLVA 176
DB 253 RLKNSETSID-----IPDFCASLQNTVAHHSKLIHFESLSQEKLPQO---LVI 302
QY 177 SGVANSFYIRRALBITLTAATCTLLCPPRRLCTDNGIMIANNGIERLRAGLILHDIEG 236
DB 303 GGVANAOYIIFGAIKSLAAHNVTTIKVLISCTDNAMEIAYSGL-----LMLVNRSEA 356
QY 237 IRYEP-----KCPLGVDISKEVGEASIKVPQLKM 265

```

Db 357 IMWRPNDIPTIYAHASDIDGTDSSEI----IDTPRKL 392

RESULT 11  
US-08-087-797-2

; Sequence 2, Application US/08087797  
; Patent No. 5543312

; GENERAL INFORMATION:

; APPLICANT: Mellors, Alan

; APPLICANT: lo, Reggie Y C.

; APPLICANT: Abdullah, Khalid M.

; TITLE OF INVENTION: Pasteurella Haemolytica

; TITLE OF INVENTION: Glycopolase

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.

; STREET: 1211 East Morehead Street,

; CITY: Charlotte

; STATE: No. 5543312ch Carolina

; COUNTRY: United States

; ZIP: 28234

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/087,797

; FILING DATE: 14-JUL-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Layton, Jr., Samuel G.

; REGISTRATION NUMBER: 22807

; REFERENCE/DOCKET NUMBER: 3374-80

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 704 377 1561

; TELEFAX: 704 334 2014

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 325 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-087-797-2

Query Match 19.5%; Score 270.5; DB 1; Length 325;

Best local Similarity 35.0%; Pred. No. 4.2e-24;

Matches 79; Conservative 26; Mismatches 102; Indels 19; Gaps 6;

QY 1 MEAHALITRL-TNKVEPFLVLLISGHCILALVGVSDFLLGKSLDIAPGMDLKVAR 59

Db 112 MEGLHAPMLEENAPPEPVALISGHTQLVKVDVGVQYELLGSEIDDAAGEAFDXTKX 171

QY 60 RLSLHPEECSTMSGKAIIEHLAKQGNRFPDIKRPPLHAKNCDFSTGLQHTVDKIIMK 119

Db 172 LGLL-DYP-----AGAVMSKLAESGTPNRPKPRPMTDRPGLDFSPGLKTPAANTKA 224

QY 120 KEKEEGIEKQILSSADIAATVQHTMACLVKTRTHAILFCRKORDLLPQNNAVLVASCG 179

Db 225 NUNENGEIDBQ---TKCDIAHAFOQAV-----VDITILIKCK-RALEQGYKRLVMAAG 273

QY 180 VASNFYIRALAILTNATQCTLLCPPRLCTDNGIMIAMNGIERLAR 225

Db 274 VSANKQLRADLAEMMKLKGSVFYPFPQFCTDNGAMIAVGYFLTKX 319

RESULT 12

US-03-403-039A-9221

; Sequence 9221, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; PRIOR FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; NUMBER OF SEQ ID NOS: 14342

; LENGTH: 343

; TYPE: PRT

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-9221

Query Match 19.4%; Score 268.5; DB 4; Length 343;

Best local Similarity 32.7%; Pred. No. 8.1e-24;

Matches 81; Conservative 31; Mismatches 105; Indels 31; Gaps 8;

QY 1 MEAHALITRL-TNKVEPFLVLLISGHCILALVGVSDFLLGKSLDIAPGMDLKVAR 59

Db 118 MEGLHAPMLEDNPPAPFPVALISGHTQLISVIGGYELLGSEIDDAAGEAFDXTKX 177

QY 60 RLSLHPEECSTMSGKAIIEHLAKQGNRFPDIKRPPLHAKNCDFSTGLQHTVDKIIMK 119

Db 178 LGLL-DYP-----GGPMLSKVASQGTGGRFVFPFRPMTDRPGLDFSPGLKTPAANTIRS 230

QY 120 KEKEEGIEKQILSSADIAATVQHTMACLVKTRTHAILFCRKORDLLPQNN-ATLVASG 178

Db 231 NGDDE-----QTRADIAAFEDAVVDTLMIKCRA-----LEQGFKELVWAG 273

QY 179 GVSANFYIRALAILTNATQCTLLCPPRLCTDNGIMIAMNGIERLAGLILHIE-GI 237

Db 274 GVSANRTIRAKLAEMQGRGGEVFPYARBPFCITDNGAMIAVGMVLTQGA-----KAEIGV 329

QY 238 RYEPKCP 245

Db 330 TVRPWP 337

RESULT 13

US-09-107-532A-6609

; Sequence 6609, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,532A

; FILING DATE: 30-Jun-1998

; PRIOR APPLICATION NUMBER: 60/085,598

; FILING DATE: 14 May 1998

; APPLICATION NUMBER: 60/051571

; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Arinello, Pamela Deneke

; REGISTRATION NUMBER: 40,489

; REFERENCE/DOCKET NUMBER: GTC-012

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6609:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 363 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHEICAL: YES
;   ORIGINAL SOURCE:
;     ORGANISM: Enterococcus faecium
;   FEATURE:
;     NAME/KEY: misc feature
;     LOCATION: (B) LOCATION 1...363
;     SEQUENCE DESCRIPTION: SEQ ID NO: 6609:
US-09-107-532A-6609

Query Match      18.8%; Score 261; DB 4; Length 363;
Best Local Similarity 33.8%; Pred. No. 7.3e-23;
Matches 75; Conservative 37; Mismatches 78; Indels 32; Gaps 9;

QY 1 MEAAHATRLTNKVEPFLVLLISGGHCLALVGVSDPFLIGKSLDAPGMDPKVRR 60
DB 140 MACHIAAARLVKPFQFLMALVSGHTELVYQEDSGYEIIGTRDAAVEAYDKVGRV 199
QY 61 LSLIKPECSYSGKAIENHAKOG-NRPFDPKPLHAKNCDFSGTGLQHTVDKIMK 119
DB 200 LGL-----SYSGKEIDQLAHQKDNHF--PRAMIHEDNYPSFGSKAFINLVHN 250
QY 120 -KEKEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCQKRDLPONNAV--LVA 176
DB 251 AQRGEGLDN-----DLASFQASVIDVLIKTIRA---C-----QNVYKQDLVV 293
QY 177 SGVANSFYIRRAELITNA--TQCTLLCPPRLCTDNGIMT 216
DB 294 AGGVAAVQGRREGIQAALSAKLPEVEVLVTPRLCGDNAAVI 335

RESULT 14
US-08-987-121A-4
; Sequence 4, Application US/08987121A
; GENERAL INFORMATION:
; APPLICANT: Hoskins, Jo Ann
; APPLICANT: Tang, Joseph Chiu-Chung
; APPLICANT: Treadway, Patci Jean
; TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence
; TITLE OF INVENTION: Gcp
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/987,121A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
```

```

; LENGTH: 336 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-987-121A-4

Query Match      18.3%; Score 254; DB 3; Length 336;
Best Local Similarity 32.4%; Pred. No. 4.6e-22;
Matches 73; Conservative 36; Mismatches 86; Indels 30; Gaps 8;

QY 1 MEAAHATRLTNKVEPFLVLLISGGHCLALVGVSDPFLIGKSLDAPGMDPKVRR 60
DB 115 MACHIAAARLVKPFQFLMALVSGHTELVYQEDSGYEIIGTRDAAVEAYDKVGRV 174
QY 61 LSLIKPECSYSGKAIENHAKOG-NRPFDPKPLHAKNCDFSGTGLQHTVDKIMK 120
DB 175 MGL-----TYPAGREIDELAHQCHD-YDPPRAMIENEDNEFSFGSKAFINLVHNA 226
QY 121 EKEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCQKRDLPONNAV--LVA 180
DB 227 E-----QKGSLSLT-EDLCASFQAAVMDIIMAKTKKAL-----EKYPVK--TLVAVAGV 272
QY 181 ASNFYIRRAELITNAQCTLLCPPRLCTDNGIMT-----NN 219
DB 273 AAKGLRRL--ATEITDVNVIIPRLCGDNAGMIAVASVSEWN 315

RESULT 15
US-08-961-083-52
; Sequence 52, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: streptococcus pneumoniae Antigens and Vaccines
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 335 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-961-083-52

Query Match      18.1%; Score 250; DB 3; Length 335;
Best Local Similarity 32.0%; Pred. No. 1.4e-21;
Matches 72; Conservative 36; Mismatches 87; Indels 30; Gaps 8;
```

```

Qy 1 MEAHALTITRLTNKVEFPFLVLLISGGHCLLAVGVSPFLIGKSLDIAPGMLDKVAR 60
Db 114 MAGHIMAAQSYVEPLFPLALLVSGHTEL VVSEAGDYKI VGETRDDAVGEAYDKVGRV 173
Qy 61 LSLIKHPECSMTSGSKAIEHLAKOGRPHFDIKPPLHAKNCDPSFTGLQHTDKIMCK 120
Db 174 MGL-----TYPAGREIDELAHQODI-YDPPRAMIKEDNLEFSFSGLKSAPINLHNA 225
Qy 121 EKEGIEKQIILSSNADIATVQHTMACHLVKTRAILFCQRDLDPONNAVLVASGV 180
Db 226 E-----QKESLST-EDLCASFQAAMVDIIMAKTKKAL-----EKYPVK--ILVVAGSV 271
Qy 181 ASNFYIRRALEILTNAQCTLLCPPRLCTDNGIMIA-----WN 219
Db 272 AANKGLRERL--AAETDVKVIIPPLRLCGDNAGMIAAYASVSXWN 314

```

Search completed: November 10, 2005, 16:54:34  
 Job time : 21.4761 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: November 10, 2005, 16:48:13 ; Search time 601.911 Seconds  
(without alignment)  
2625.922 Million cell updates/sec

Title: us-10-649-273-2\_COPY\_148\_414  
Perfect score: 1385  
Sequence: 1 MEAHALITRLTNKVPEPLV.....DISKEVGASIKVQLKKEI 267

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 4390206 seqs, 2959870667 residues  
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPRO.spool\_p/US10649273/runat\_02112005\_091337\_15540/app.query.fasta\_1.1429  
-DB=N\_Geneseq\_16Dec04 -QPM=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pco -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USPR=US10649273 @CGN 1.1 1063 @runat\_02112005\_091337\_15540 -NCPU=6 -ICPR=3  
-NO MMAR -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: Geneseq\_16Dec04:\*  
2: Geneseq\_1980s:\*  
3: Geneseq\_1990s:\*  
4: Geneseq\_2000s:\*  
5: Geneseq\_2001as:\*  
6: Geneseq\_2001bs:\*  
7: Geneseq\_2002as:\*  
8: Geneseq\_2002bs:\*  
9: Geneseq\_2003as:\*  
10: Geneseq\_2003bs:\*  
11: Geneseq\_2003cs:\*  
12: Geneseq\_2004as:\*  
13: Geneseq\_2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	1385	100.0	1416	ABX70950 Novel hum
2	1395	100.0	1526	ABs76635 DNA encod
3	1385	100.0	2058	ABA93268 Human O-s
4	1385	100.0	2197	ABs76635 DNA encod
5	1385	100.0	2572	ABT23207 Human pro

6	1348	97.3	1820	AA046856	Ad46856 Human gly
7	1348	97.3	1820	ACA08887	ACA08887 Human cDN
8	1348	97.3	1821	ABs57020	ABs57020 cDNA encod
9	1204	86.9	2208	ADA52832	ADA52832 Human cod
10	1204	86.9	2890	ADQ24627	Adq24627 Human sof
11	995.5	71.9	3358	ADs31345	Ad31345 Human dia
12	983	71.0	1572	AB075508	AB075508 Murine si
13	599	43.2	2734	5 AAs84622	AAs84622 DNA encod
14	468	33.8	371	ADL86725	ADL86725 DNA up-re
15	468	33.8	371	ADL86726	ADL86726 DNA up-re
16	415.5	30.0	1601	AB124633	AB124633 Drosophi1
17	415.5	30.0	3656	AB124632	AB124632 Drosophi1
18	401.5	29.0	1557	3 AAC38454	AAC38454 Arabidops
19	342	24.7	4360	6 AAD48239	AAH15110 Human cDN
20	338.5	24.4	1385	4 AAH15110	AAH15110 Human cDN
21	321.5	23.2	1146	8 ACA26804	ACA26804 Prokaryot
22	314.5	22.7	936	8 ACA20445	ACA20445 Prokaryot
23	308.5	22.3	1044	8 ACA39102	ACA39102 Prokaryot
24	308.5	22.3	94750	4 AAF28551	AAF28551 Genomic f
25	307	22.2	1053	12 ADL03120	ADL03120 DNA encod
26	304.5	22.0	1032	8 ACA43173	ACA43173 Prokaryot
27	301.5	21.8	1000	4 AAF91424	AAF91424 Moraxella
28	301.5	21.8	1000	6 ABQ37804	ABQ37804 DNA seque
29	300	21.7	1092	6 ABQ90383	ABQ90383 M. capsul
30	291.5	21.0	1026	4 AAs54064	AAs54064 Pseudomon
31	291.5	21.0	1026	8 ACA42146	ACA42146 Pseudomon
32	291.5	21.0	1026	10 ADG73341	ADG73341 P aerugin
33	291.5	21.0	1059	11 ABD02280	ABD02280 Pseudomon
34	291.5	21.0	1206	11 ABD02197	ABD02197 P aerugin
35	289.5	20.9	1026	10 ADG73343	ADG73343 P aerugin
36	289.5	20.9	1029	4 AAs53309	AAs53309 Haemophil
37	289.5	20.9	1029	8 ACA34150	ACA34150 Haemophil
38	289.5	20.9	110000	2 AAT42063_05	Continuation (6 of
39	285	20.6	372	5 ABV61069	ABV61069 Human pro
40	282.5	20.4	9667	13 ADT05493	ADT05493 Haemophil
41	282.5	20.4	85814	13 ADT05644	ADT05644 Haemophil
42	280.5	20.3	1020	8 ACA44384	ACA44384 Prokaryot
43	280.5	20.3	1074	10 ADF02056	ADF02056 Bacterial
44	279.5	20.2	1014	4 AAs56045	AAs56045 Salmone11
45	279.5	20.2	1014	8 ACA51431	ACA51431 Prokaryot

## ALIGNMENTS

RESULT 1	ABX70950	
ID	ABX70950 standard; cDNA, 1416 BP.	
XX		
AC	ABX70950;	
XX		
DT	05-MAR-2003 (first entry)	
XX		
DE	Novel human cDNA sequence #175.	
XX		
KW	Human; gene; ss; nervous system disorder; peripheral neuropathy;	
KW	Huntington's disease; amyotrophic lateral sclerosis; hemophilia;	
KW	neurodegenerative disease; Parkinson's disease; Alzheimer's disease;	
KW	autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;	
KW	insulin-dependent diabetes mellitus; anaemia; thrombocytopenia; wound;	
KW	ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;	
KW	fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;	
KW	coagulation disorder; cancer; tumour; inflammatory disease; septic shock;	
KW	Crohn's disease; anaplasia; proliferation; haematopoiesis; chemokineic;	
KW	differentiation; stem cell growth factor; haematopoiesis; chemokineic;	
KW	haemostatic; antiinflammatory; expressed sequence tag; EST.	
OS	Homo sapiens.	
XX		
PN	W0200281731-A2.	
XX		
PD	17-OCT-2002.	
XX		
PF	29-JUN-2002; 2002WO-US001222.	

XX 30-JAN-2001; 2001US-00774528.  
 PR (HYSE-) HYSEO INC.  
 PA (GOOD/) GOODRICH R W.  
 XX  
 XX Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xie AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;  
 XX WPI; 2003-058563/05.  
 DR  
 XX Novel polypeptide useful for treating neurodegenerative diseases, myeloid  
 PT or lymphoid cell disorders, bone disorders, mechanical and traumatic  
 PT disorders, coagulation disorders, and inflammatory diseases.  
 XX  
 PS Claim 1; Page: 612pp; English.

CC This invention relates to the cDNA sequences encoding an isolated novel  
 CC human polypeptide. The protein encoded by the nucleic acid of the  
 CC invention is useful for treating central and peripheral nervous system  
 CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic  
 CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,  
 CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus  
 CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)  
 CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopenia)  
 CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,  
 CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head  
 CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;  
 CC bacterial, viral or fungal infections; allergic conditions such as  
 CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);  
 CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's  
 CC disease, anaphylaxis). The protein may be used to inhibit the growth,  
 CC infection or function of infectious agents such as bacteria, fungi,  
 CC viruses, or to effect bodily characteristics, biorhythms or circadian  
 CC cycles of rhythms. The protein may also have  
 CC proliferation/differentiation, stem cell growth factor, haematopoiesis  
 CC regulation, immune stimulation or suppressing, chemotactic/chemokinetic,  
 CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory  
 CC activities. The cDNA sequences of the invention are useful for expressing  
 CC recombinant protein for analysis. The present sequence represents a novel  
 CC human cDNA sequence of the invention, this sequence is an expressed  
 CC sequence tag (EST) and was identified using subtractive hybridisation  
 XX  
 SO Sequence 1416 BP; 441 A; 280 C; 273 G; 422 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 4.68e-147 Length: 1416  
 Score: 1385.00 Matches: 267  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 8 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x ABX70950 (1-1416)

QY 1 MetGUAUAAHSAaleuthrTleargleuthrAenlvsvaIGlupherProheleuVal 20  
 Db 502 ATGGAGGCTCAATGACCTTACTTATGTAGGTTGACCAATTAAGAAATTTCTTTTAAAGTT 561  
 QY 21 LeuLeuIleSerGlyGlyHisCysleuLeuAlaLeuValGlyValSerAspPheLeu 40  
 Db 562 CTTTGGATTTCTGGAGGCTACGCTGCTGGCATTAAGTTCAAGAGCTTTCAAGATTTCTG 621  
 QY 41 LeuLeuGlyIleSerIleuAspIleAlaProGlyAspMetIleuAspIleValAlaArgArg 60  
 Db 622 CTTCTTGAAGAGCTTTTGGACATAGACCAAGTGCATGCTTGACAAAGTGGCAAGAAG 681  
 QY 61 LeuSerIleuIleIleHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
 Db 682 CTTTCTTAATTAACAATCCAGAGTGCTCCACCATAGAGTGAGGAAAGCATAGAAACAT 741  
 QY 81 LeuAlaIleGlyGlyValAsnArgPheHisPheAspIleIleValProProIleuHisAlaIleVal 100

Db 742 TTGGCCAAACAGAGAAATGATTTGATTTGACATCAAACTCCCTGCATCATGCTAAA 801  
 QY 101 AenCysAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleIleMetIleVal 120  
 Db 802 AATTGTGATTTTCTTTTACTGAGCTTCAACGTTACTGATTAATTAATTAATGAAG 861  
 QY 121 GluIleGlyGlyGlyIleGlyIleGlyIleGlyIleLeuSerSerAlaAlaAspIleAla 140  
 Db 862 GAAAAAGAGAGATTTAGAGAGGGCAAACTCTCTTTCAGACAGACATTCCTGCC 921  
 QY 141 ThrValGlnHisThrMetAlaCysHisIleuValIleValArgThrHisArgAlaIleLeuPhe 160  
 Db 922 AAGATACAGACACACATGCGATGCTTGTGAAAAAGAACATCGGCTATTCGT 961  
 QY 161 CysIleGlnArgAspIleuProGlnAsnAsnAlaValIleuValAlaSerGlyIleVal 180  
 Db 982 TGTAGCAGAGAGACTTGTACTCTCAAAATTAATGAGTACTGTTGCATCTGTTGCTG 1041  
 QY 181 AlaSerAsnPheThrIleArgArgAlaLeuGlyIleLeuThrAsnAlaThrGlnCysThr 200  
 Db 1042 GGAAGTAACTTCTATATCCGAGAGCTTGAAATTTTAACAAACGCAACACAGTCACT 1101  
 QY 201 LeuLeuCysProProProArgIleuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
 Db 1102 TTTGTGTGTCCTCTCCAGACTATGCACTGATTAATGGAATTAATGATGATGAAATG 1161  
 QY 221 IleGluArgLeuArgAlaGlyIleuGlyIleLeuHisAspIleGlyIleArgIleVal 240  
 Db 1162 ATTTGAAGACTGAGTCTGCTGGCATTTTACATGACATGAGAGCATCCGCTATGAA 1221  
 QY 241 ProIleCysProLeuGlyValAspIleSerIleGlyValGlyIleAspIleVal 260  
 Db 1222 CCAAAATGCTCTTGTGAGTATGACATATCAAAAGAAATTTGAGAACTTCATTAAGA 1281  
 QY 261 ProGlnLeuIleMetGlyIle 267  
 Db 1282 CCAATTAATAAATGAGAGATA 1302  
 AC  
 XX  
 DT 11-DEC-2002 (first entry)  
 DE  
 XX  
 DE DNA encoding novel human metalloprotease MPl fragment #1.  
 XX  
 XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
 KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
 KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
 KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
 KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
 KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
 KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
 KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
 KW neurological disorder; gene; ds.  
 KW  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200272751-A2.  
 XX  
 PD 19-SEP-2002.  
 XX  
 PF 05-FEB-2002; 2002MO-US003353.  
 XX  
 PR 05-FEB-2001; 2001US-0266518P.  
 PR 10-APR-2001; 2001US-0282814P.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;



XX MPI; 2002-723329/78.  
DR F-PSDB; ABG96487.

XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
PT treating, or ameliorating diseases associated with aberrant  
PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and  
PT neurological disorders.

XX Disclosure; Page 462-463; 473pp; English.

XX The invention describes an isolated nucleic acid molecule (I) encoding a  
CC metalloproteinase (MP-1). (I) is useful for preventing, treating, or  
CC ameliorating a medical condition, particularly an immune disorder, an  
CC aberrant glutamate transport or motor neuron disorder, such as  
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
CC condition. The compositions and methods are also useful for diagnosing,  
CC prognosticating, treating, ameliorating and/or treating disorders  
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease). This sequence represents a  
CC metalloproteinase MP1 polynucleotide  
XX

XX Sequence 1526 BP; 484 A; 297 C; 289 G; 456 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 5.2e-147 Length: 1526  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x ABS76639 (1-1526)

QY 1 MetGluAlaHisAlaLeuThr11eArgLeuThrAsnValGluPheProPheLeuVal 20  
DB 1 ATGGAGGCTCATGCACTTACTATTAGGTTGACCAATAGAGTAATTCCTTTTATGTT 60  
QY 21 LeuLeu11SerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 61 CTTTGAATTTCTGAGGTCACTGCTGTTGGCATTAAGTTCAGAGATTTCTG 120  
QY 41 LeuLeuGlyLysSerLeuAsp11eAlaProGlyAsnMetLeuAspLysValAlaArgArg 60  
DB 121 CTTCTTGGAAGCTTTGGACATAGCACAGGTACATCTCTTGACAGGTGGCAAGAGA 180  
QY 61 LeuSerLeu11eLysHisProGluCysSerThrMetSerGlyGlyValAla11eGluHis 80  
DB 181 CTTTCTTTAATAAATCATCCAGAGTCTCCACCATGATGATGGTGAAGCATGAACAT 240  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAsp11eLysProPoleuHisHisAlaLys 100  
DB 241 TTGGCCAAACAGGAATATGATTCATTTGACATCAACTCCCTTGCCATCATGCTAAA 300  
QY 101 AaaCysAspPheSerPheThrGlyLeuGlnHisValThrAspLys11eMetLysLys 120  
DB 301 AATGTGATTTTCTTTACTGAGCTTCAACAGCTTACTGTAATAATATATATAAAG 360  
QY 121 GlnLysGlnGluGly11eGlnLysGlyGln11eLeuSerSerAlaAlaAsp11eAlaAla 140  
DB 361 GAAAAAGAGAGATATGAGAAAGGGCAATCTCTGCTTACAGACAGACATTTGCTCC 420  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAla11eLeuPhe 160  
DB 421 ACAGTACAGACACAAATGCGATGTCATCTGTGAAAGAAACATCGGGCTATTCTGTT 480

QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
DB 481 TGTAAAGCAGAGAGACTTGTACTCTAAATATGACAGTACGTGTCATCTGTGGTGC 540  
QY 181 AlaSerAsnPheThr11eArgArgAlaLeuGlu11eLeuThrAsnAlaThrGlnCysThr 200  
DB 541 GCAAGTACTTTATATCCGACAGGCTCTGGAAATTTTACAAACCAACACAGTGCAT 600  
QY 201 LeuLeuCysProProProArgLeuGlyCysThrAspAsnGly11eMet11eAlaTPaenGly 220  
DB 601 TTGTTGTGTCCTCTCCACGACATGACATGATATGATGATTCATGCAATGGG 660  
QY 221 11eGlnLysGlnArgAla11eGlyGlyGly11eLeuHisAsp11eGlnGly11eArgGly 240  
DB 661 ATTTGAAGACTACGCTGCTGGCATTTTAAATGAAAGCATATCCGCTATGAA 720  
QY 241 ProLysCysProLeuGlyValAsp11eSerLysGlnValGlyGlnAlaSer11eLysVal 260  
DB 721 CCAAAATGCTCTTGGAGTACATATCAAAAGAAATTTGAGAGACCTTCATAAAGTA 780  
QY 261 ProGlnLeuLysMetGlu11e 267  
DB 781 CCACATTTAAATGAGATA 801  
RESULT 3  
ABA93268  
ID ABA93268 standard; cDNA; 2058 BP.  
AC ABA93268;  
DT 19-APR-2002 (first entry)  
XX  
XX  
DE Human O-sialoglycoproteinase-like protein encoding cDNA SEQ ID NO:1.  
XX  
XX Human; O-sialoglycoproteinase-like protein; OSGPLP; enzyme; gene; ss.  
XX Homo sapiens.  
OS  
FH Key Location/Qualifiers  
FT CDS 110..1354  
FT /tag= a  
FT /product= "O-sialoglycoproteinase-like protein"  
XX  
XX CNA138550-A.  
XX  
XX 24-OCT-2001.  
XX  
XX 19-APR-2000; 2000CN-00106834.  
XX PF  
XX 19-APR-2000; 2000CN-00106834.  
XX PR  
XX (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.  
XX  
XX Mao Y, Xie Y;  
PI MPI; 2002-115090/16.  
XX DR  
XX P-PSDB; ABB05481.  
XX  
XX  
XX O-sialoglycoproteinase-like protein and encoding polynucleotide, useful  
PT for diagnosing, preventing and treating related diseases.  
XX  
XX  
XX Claim 5; Page 29-30 (Disclosure); 38pp; Chinese.  
PS  
XX  
XX The present sequence encodes human O-sialoglycoproteinase-like protein  
CC (OSGPLP). The present invention also describes: (1) the preparation of  
CC the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the  
CC prevention and/or treatment of related diseases; (4) utilizing the OSGPLP  
CC protein in screening its agonist, excitomotor and inhibitor and preparing  
CC an antibody against the OSGPLP protein; and (5) the use of the OSGPLP  
CC polynucleotide sequences, proteins, agonists, excitomotors, inhibitors  
CC and antibodies in treating diseases related to the abnormal OSGPLP gene  
CC and in preparing the medicine composite for the treatment

SQ Sequence 2058 BP; 637 A; 400 C; 410 G; 611 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	7,97e-147	Length:	2058
Score:	1385.00	Matches:	267
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2\_COPY\_148\_414 (1-267) x ABS93268 (1-2058)

```
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
Db 551 ATGGAGGCTCATGCATCTTACTATTAGTTGACCATTAAGATGATTTCTTTTATTAGTT 610
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 611 CTTTGTGATTTCTGGAGGCTGCTGCTGTCGATTTAGTTCAAGAGTTTCAGATTTCTCG 670
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db 671 CTTCTTGGAAAGCTTTGGACATAGCACAGGTGACATGCTTGACAGGTGGCAAGAAAGA 730
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
Db 731 CTTTCTTTAATTAACATCCAGAGTCTCCACCATGATGGTGGGAAAGCCATAGAACAT 790
QY 81 LeuAlaLysGlnGlyAsnArgPheHisAspIleLysProProLeuHisAlaLys 100
Db 791 TTGGCCAAACAGGAAGAAATGATTTGATTTTGGATCAATCCCTTCGATCATGCTAAA 850
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValIleThrAspLysIleIleMetLysLys 120
Db 101 AATTGTGATTTTCTTTTACTGAGCTTCAACACGTTACTGATTAATAATATGAAAAAG 910
QY 121 GluLysGluGluGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIa 140
Db 911 GAAABAAGGAAAGGATTTGAGAAAGGGCAATCTGCTTCGACGACGACATTCCTGCC 970
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
Db 971 ACAGTACAGCACACATGCGATGCTCTTGGAAGAAAGAACATCGGGCTATTCGTTT 1030
QY 161 CysLysGlnIleArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180
Db 1031 TGTAAAGCAGAGAGACTGTGTTACTCCTCAAAATATATGACAGTACTGTGTGTGTC 1090
QY 181 AlaSerAsnPheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
Db 1091 GCAAGTAACTTCTATATCCGAGAGCTCTGGAAATTTTAAACAAAGCCACACAGTGCCT 1150
QY 201 LeuLeuCysProProCysProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
Db 1151 TTGTGTGTCTCTCCGCCAGACTATGCACTGATTAATGATATGATTCATGAGATGCT 1210
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrlu 240
Db 1211 ATTGAAGAAGCTACGCTGCTGGCTTGGCATTTTAACTGATGAAAGGATCCGCTATGAA 1270
QY 241 ProLysCysProLeuGlyValAspIleSerLysGlyValGlyAlaSerIleLysVal 260
Db 1271 CCAAAATGTCCTCTTGGAGTGAACATATCAAAAGAAAGTTGGAAAGCTTCATAAAGTA 1330
QY 261 ProGlnLeuLysMetGluIle 267
Db 1331 CCAACATTAATAATGAGATTA 1351
```

RESULT 4  
ABS76635  
ID ABS76635 standard; DNA; 2197 BP.  
XX  
AC ABS76635;

XX 11-DEC-2002 (first entry)

DE DNA encoding novel human metalloprotease MPI.

KW Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
KW emphysema; cystic fibrosis; inflammatory disorder; inflammatory disorder;  
KW neurological disorder; gene; ds.

OS Homo sapiens.

PN WO200272751-A2.

XX 19-SEP-2002.

PF 05-FEB-2002; 2002MO-US003353.

XX 05-FEB-2001; 2001US-0266518P.

PR 10-APR-2001; 2001US-0282814P.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;

PI MPI; 2002-723329/78.

XX P-PSDB; ABS66478.

PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
PT creating, or ameliorating diseases associated with aberrant  
PT metalloprotease activity, e.g. immune, metabolic, inflammatory and  
PT neurological disorders.

XX Claim 1, Fig 1A-C; 473pp; English.

XX The invention describes an isolated nucleic acid molecule (I) encoding a  
CC metalloprotease (MP-1). (I) is useful for preventing, treating, or  
CC ameliorating a medical condition, particularly an immune disorder, an  
CC aberrant glutamate transport or motor neuron disorder, such as  
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
CC condition. The compositions and methods are also useful for diagnosing,  
CC prognosticating, treating, ameliorating and/or treating disorders  
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease). This sequence represents a  
XX metalloprotease MPI polynucleotide

SQ Sequence 2197 BP; 681 A; 441 C; 439 G; 636 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	8,75e-147	Length:	2197
Score:	1385.00	Matches:	267
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2\_COPY\_148\_414 (1-267) x ABS76635 (1-2197)

```
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
|||||
```

Db 672 ATGAGGCTCATGCACTTACTATTAGTTCAGCAATAAGTAGAATTCCTTTTACTT 731  
 QY 21 LeuLeuIleSerGlyGlyYhiCysLeuLeuAlaLeuValGlnGlyValSerAspHeu 40  
 Db 732 CTTTGAATTTCTGAGAGTCACTGCTGTTGGCATTAGTTCAGAGATTTCAGATTTTCTG 791  
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
 Db 792 CTTCTTGAAAGTCTTTGGACATGACACAGGTGCATGCTTGACAAAGTGCAGAAAGA 851  
 QY 61 LeuSerLeuIleLysNHISpGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
 Db 852 CTTTCTTTAATAAATCATCCAGAGTCTCCACATGAGTGGTGGAAAGCCATAGAACAT 911  
 QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
 Db 912 TTGGCCAAACAAAGAAATAGATTTTCATTTTGACATCAAACTCCCTTGATCATGCTAA 971  
 QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
 Db 972 AATTGTATTTTCTTTTACTGACCTTCACACGTTACTGATTAATAATATGAAAAAG 1031  
 QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
 Db 1032 GAAAAAGGAGGATTTGAGAGGGGCAAAATCCGTCTTACGACGACATTCCTGCTCC 1091  
 QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
 Db 1092 ACGATACAGCACACAAATGCGATGTCATCTTGTGAAAAAACAACATCGGGCTATTCGTTT 1151  
 QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValAlaLeuValAlaSerGlyGlyVal 180  
 Db 1152 TGTAAAGAGAGAGACTTGTACTCAAAATATATGACATGCTGTCATCTGCTGCTGC 1211  
 QY 181 AlaSerAsnPheIleYllLeuArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
 Db 1212 GCAAGTAACTCTCTATATCCGAGAGCTCTGGAATTTTAAACAAACAGACAGTCACT 1271  
 QY 201 LeuLeuCysProProArgLysCysThrAspAsnGlyIleMetIleAlaIleArgAsnGly 220  
 Db 1272 TTGTGTGCTCTCTCCAGACTATGACATATATATGAGCATTTATGATTCATGAGATGCT 1331  
 QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240  
 Db 1332 ATTGAAAGACTACCTGCTGCGCTTGGGCAATTTTACATGACATAGAAAGCATCCGATGGA 1391  
 QY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260  
 Db 1392 CCAAAAATGCTCTCTTGAGATGACATATCAAAAGAACTTGGAAGCTTCCATAAAGTA 1451  
 QY 261 ProGlnLeuLysMetGluIle 267  
 Db 1452 CCACAATTAATAAATGAGATTA 1472  
 Db 1452 CCACAATTAATAAATGAGATTA 1472  
 RESULT 5  
 ID ABT23207 standard; DNA; 2572 BP.  
 AC ABT23207;  
 XX 01-MAY-2003 (first entry)  
 XX Human protein modification + maintenance molecule DNA SEQ ID No 36.  
 XX  
 XX Cytosolic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;  
 KM cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;  
 KM antitumor; hepatotropic; gynaecological; antibacterial; vincide;  
 KM protozoicide; antiparasitic; cell proliferative disease; PMOD;  
 KM protein modification and maintenance molecule; immunogenic fragment;  
 KM cancer; autoimmunity; inflammatory disease; neurological disorder;  
 KM gastroenteric; developmental; vesicle trafficking disorder; infection;  
 KM protein-protein interaction; drug-target interaction;  
 KM gene expression profile; human; gene; de.

XX OS Homo sapiens.  
 XX XX WO200300844-A2.  
 XX XX  
 PD 03-JAN-2003.  
 XX XX  
 PF 18-JUN-2002; 2002MO-US019360.  
 XX XX  
 PR 22-JUN-2001; 2001US-0300508P.  
 PR 06-JUL-2001; 2001US-0303445P.  
 PR 13-JUL-2001; 2001US-0305405P.  
 PR 09-AUG-2001; 2001US-0311442P.  
 PR 24-AUG-2001; 2001US-0314821P.  
 PR 29-AUG-2001; 2001US-0315992P.  
 PR 03-MAY-2002; 2002US-0378205P.  
 XX XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX XX  
 PI Gandhi AR, Kable AE, Swarnakar A, Hafalia AUA, Tran B, Dugan BM,  
 PI Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DAM, Lee EA, Yue H;  
 PI Foraythe JF, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J;  
 PI Thangavelu K, Gietzen KJ, Ding L, Baughn MR, Borowsky ML, Yao MG;  
 PI Wallia NK, Mason PM, Gururajan R, Lee S, Becha SD, Lee ST, Tran UK;  
 PI Elliott VS, Luo W, Sprague MW, Tang YT, Lu Y, Zebardjian Y;  
 XX XX  
 DR WPI; 2003-184039/18.  
 DR P-PSDB; ABJ26654.  
 XX XX  
 PT New isolated human PMOD polypeptide and polynucleotide, useful for  
 PT diagnosing, treating and preventing diseases or conditions associated  
 PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and  
 PT infections.  
 XX XX  
 PS Claim 91; Page 211; 225pp; English.  
 XX XX  
 CC The invention relates to an isolated polypeptide comprising: any of 28  
 CC sequences of 48-1256 amino acids; a natural amino acid sequence at least  
 CC 90% identical to the 28 amino acid sequences; 94% identical to a sequence  
 CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino  
 CC acids, or 97% identical to a sequence of 242 amino acids, all given in  
 CC the specification; or a biologically active or immunogenic fragment of  
 CC the isolated polypeptide. The polypeptides and polynucleotides are useful  
 CC in diagnosing, treating and preventing diseases or conditions associated  
 CC with the decreased expression of protein modification and maintenance  
 CC molecules (PMOD), such as cell proliferative diseases (e.g. cancer,  
 CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,  
 CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,  
 CC epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.  
 CC endometriosis), developmental, vesicle trafficking disorders, and  
 CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also  
 CC useful in assessing the effects of exogenous compounds on the expression  
 CC of nucleic acid and amino acid sequences of PMOD. The PMOD or its  
 CC fragments are useful in screening compounds for effectiveness as agonist  
 CC or antagonist of the polypeptides, or in altering the expression of the  
 CC target polynucleotide and compounds that specifically bind to or modulate  
 CC the activity of the polypeptide. The microarray is useful in monitoring  
 CC or measuring protein-protein interactions, drug-target interactions, and  
 CC gene expression profiles. This polynucleotide sequence represents the DNA  
 CC encoding a human PMOD protein of the invention  
 XX XX  
 SQ Sequence 2572 BP; 780 A; 489 C; 509 G; 794 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 1,09e-146 Length: 2572  
 Score: 1385.00 Matches: 267  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 8 Gaps: 0  
 US-10-649-273-2\_COPY\_148\_414 (1-267) x ABT23207 (1-2572)



Db 647 CTTTGGATTTCGAGGCTCAGCTGCTGTTGGCATTGATTCAGAGGATTTCAGATTTTCG 706  
Qy 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
Db 707 CTTCTTGGAAAGTCTTTGGACATAGCACAGGTGACATCTTGGACAGGTGGCAAGAGA 766  
Qy 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysAlaIleGluHis 80  
Db 767 CTTTCTTTAATAAACAATCCAGAGTCTCCACCATGAGGTGGGAAAGCCATTAGAACAT 826  
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
Db 827 TTGGCCAAACAGAAATAGATTTCATTGATTCACCAACCTCCCTTGATCATGCTAA 886  
Qy 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
Db 887 AATTGTGATTTTCTTTTACTGACTTCAACACGTTACTGATMAAATATATGAAACAG 946  
Qy 121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
Db 947 AAACAAAGGAGGATATGAGAGGGCAAAATCTGTCTTGACGACGACATTCCTGCC 1006  
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgHisArgAlaIleLeuPhe 160  
Db 1007 ACAGTACAGCACACAATGGCATGTCTGTTGAAAAGACACATCGGCTATTCTGTTT 1066  
Qy 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleValAlaSerGlyVal 180  
Db 1067 TGTAGCAGAGAGCTGTGTAACCTCAAAATATAGAGTACGTGTGCAATCTGTGTGTC 1126  
Qy 181 AlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
Db 1127 GCAAGTAACTTCTATATCCGAGAGCTCTGGAAATTTTACAAAGCAACACAGTCACT 1186  
Qy 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
Db 1187 TTGTGTGTCTCTCCACAGACTATGACATGATATGACATTAATGATTCAGATGATGCT 1246  
Qy 221 IleGluLysArgLeuAspAlaGlyLeuGlyLysIleLeuHisAspIleGluGlyIleArgTyrGlu 240  
Db 1247 ATTTAAAGACTACGTGTCTGGCTTGGCATTTTATCATGACATAGAGGATCCGCTATGAA 1306  
Qy 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260  
Db 1307 CCAAAAGTCTCTTGGAGTGAATATCAAAAGAGTTGGAGAGCTTCCATTAAGTA 1366  
Qy 261 ProGlnLeuLysMetGluIle 267  
Db 1367 CCACATTTAAAAATGAGATA 1387

RESULT 7  
ACA60887  
ID ACA60887 standard; cDNA, 1820 BP.  
XX  
AC ACA60887;  
XX  
DT 08-JUL-2003 (first entry)  
XX  
DE Human cDNA 28472 encoding a glycoprotease.  
XX  
KW Human; ss; gene; cancer; aberrant cellular proliferation;  
KW differentiation; immune disorders; heart disorder; brain disorder;  
KW cardiovascular disorder; endothelial cell disorder; pain disorder;  
KW haemopoietic disorder; blood vessel disorder; metabolic disorder;  
KW liver disorder; platelet disorder; glycoprotease.  
XX  
OS Homo sapiens.  
XX  
FT Key location/Qualifiers  
FT CDS 146..1390  
FT /\*tag= a  
FT /product= "Glycoprotease"

PT /note= "This CDS is specifically claimed in claim 1"  
XX  
PN US2003009017-A1.  
XX  
PD 09-JAN-2003.  
XX  
PF 08-NOV-2001; 2001US-00012140.  
XX  
PR 08-NOV-2000; 2000US-0246768P.  
PR 08-NOV-2000; 2000US-0246772P.  
PR 15-NOV-2000; 2000US-0249185P.  
XX  
PA (LEIB/) LEIBY K R.  
PA (KAPE/) KAPPELLER-LIBERMANN R.  
PA (GLUC/) GLUCKSMANN M A.  
PI Leiby KR, Kapellier-libermann R, Glucksmann MA;  
XX  
DR WPI, 2003-428888/40.  
DR P-PSDB; ABU09569.  
XX  
PT New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid  
PT molecules, useful for diagnosing, treating cancer, pain, or immune,  
PT heart, endothelial cell, hemopoietic, blood vessel, brain, metabolic  
PT and liver disorders.  
XX  
PS Claim 2; Fig 8; 90pp; English.  
XX  
CC The invention relates to an isolated 38650 (encoding adenosine  
CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7  
CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or  
CC a sequence which is at least 60% identical to the six nucleic acids or  
CC their open reading frames, fragments of at least 15 nucleotides,  
CC naturally occurring variants, or a DNA insert of the plasmid deposited  
CC with the American Type Culture Collection as Accession No. not defined in  
CC the specification, which encodes the amino acid sequence). Also included  
CC are a host cell containing the nucleic acids (used to produce the  
CC protein), the encoded proteins, an antibody that selectively binds to  
CC the polypeptide, and identifying a compound that binds to/modulates the  
CC activity of the polypeptide. The nucleic acid molecules, polypeptides and  
CC methods are useful for diagnosing, treating cancer, aberrant cellular  
CC proliferation and/or differentiation, immune disorders, heart disorders,  
CC cardiovascular disorders including endothelial cell disorders,  
CC haemopoietic disorders, blood vessel disorders, brain disorders, pain  
CC and metabolic disorders, liver disorders and platelet disorders (many  
CC examples of these disorders are given in the specification). The present  
CC sequence is the Human cDNA 28472 encoding a glycoprotease  
XX  
SQ Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;  
XX  
Alignment Scores:  
Pred. No.: 1,07e-142 Length: 1820  
Score: 1348.00 Matches: 260  
Percent Similarity: 98.50% Conservative: 3  
Best Local Similarity: 97.38% Mismatches: 4  
Query Match: 97.33% Indels: 0  
DB: 8 Gaps: 0  
XX  
US-10-649-273-2\_COPY\_148\_414 (1-267) x ACA60887 (1-1820)  
Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
Db 587 ATGGAGGCTCAGACCTTACTTATGAGTGGTACCAGAAATGAAATTTCTTTTATGATT 646  
Qy 21 LeuLeuLysSerGlyLysHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
Db 647 CTTTGGATTTCGAGGCTCAGCTGCTGTTGGCATTAATTCAGAGATTTTCAGATTTTCG 706  
Qy 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
Db 707 CTTCTTGGAAAGTCTTTGGACATAGCACAGGTGACATCTTGGACAGGTGGCAAGAGA 766  
Qy 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysAlaIleGluHis 80





```
QY      81 LeuAlaIaIySGInGIyAaNaArGpHeHiePheAspIleIySProProLeuHiHiaIaIyS 100
DB      828 TTGGCCCAAGAGAGAAATAGATTTCATTGACATCAAACTCCCTTCATCATCTCTAAA 887
QY      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIySileIleMetIySlyS 120
DB      888 AATTGTGATTTTCTTTTACTGAGCTTCACACGTTACTGTATTAATAATGAAGAAAGG 947
QY      121 GluIySGInGIyGIleGluIySGIyGlnIleuSerSerAlaIaAspIleAlaIa 140
DB      948 AAACAAGAGAGAGATTGAGAGAGGCAAAATCTGTCTTACGACGACGACATTCCTGCC 1007
QY      141 ThrValGlnHisThrMetAlaCysHisIleuValIySArGThrHisArGAlaIleuPhe 160
DB      1008 ACAGTACAGACACAAATGGCATGTCACTTGTGAAAAGAACACATCGGGCTATTCTGTTT 1067
QY      161 CysIySGInArGAspLeuPheProGlnAsnAsnAlaValIleuValAlaSerGIyGIyVal 180
DB      1068 TGTAAAGCAGAGAGACTGTGTAACCTCAAAATATGCAGTACGTGTCATCTGTGTGTC 1127
QY      181 AlaSerAsnPheTyrlleArGArGAlaIleuGlnIleuThrAsnAlaThrGlnCysThr 200
DB      1128 GCAAGTAACTTCTATATCCGACAGCTCGAATAATTTTAAACAACGCAACACAGTGCAC 1187
QY      201 LeuIleuCysProProArGLeuCysThrAspAsnGIyIleMetIleAlaTrpAsnGIy 220
DB      1188 TTGTGTGTCTCTCTCCACAGCTATGCATGATATGATGATGATGATGATGATGATGAT 1247
QY      221 IleGluArGLeuArGAlaGIyLeuGIyIleuHisAspIleGIyIleArGTrGIu 240
DB      1248 ATTAAGAAAGCTACGTCGTGGCTGGGCAATTTTACATGACATAGAAGCATCCGCTATGAA 1307
QY      241 ProIyCysProLeuGIyValAspIleSerIySGIyValGIyGlnAlaSerIleIySVal 260
DB      1308 CCAAAATGCTCTCTTGGAGTACATATCAAAAGAGTGGAGAGACTTCCATATAAAGTA 1367
QY      261 ProGlnIleuIySmetGIuIle 267
DB      1368 CCAATTAATAAATGAGATTA 1388
RESULT 9
AD52832
ID      AD52832 standard; cDNA; 2208 BP.
XX
AC      AD52832;
XX
DT      20-NOV-2003 (first entry)
XX
DE      Human coding sequence, SEQ ID 400.
XX
KW      Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW      Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW      inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS      Homo sapiens.
XX
PN      BP1293569-A2.
XX
PD      19-MAR-2003.
XX
PF      21-MAR-2002; 2002EP-00006586.
XX
PR      14-SEP-2001; 2001JP-00328381.
XX
PR      24-JAN-2002; 2002US-0350435P.
XX
PA      (HELI-) HELIX RES INST.
PA      (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI      Isugai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI      Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI      Seki N, Yoshikawa T, Otsuka M, Nagahari K, Maeuho Y;
```

```
DR      MPI, 2003-395539/38.
DR      - P-PSDB; AD54471.
XX
PT      New polynucleotides encoding full-length polypeptides, e.g. secretory
PT      and/or membrane proteins, useful for developing medicines for diseases in
PT      which the gene is involved, or as target molecules for gene therapy.
XX
PS      Claim 1; SEQ ID NO 400; 205pp; English.
XX
CC      The present invention relates to novel human secretory or membrane
CC      proteins (AD54072-AD55710) and their coding sequences (AD54243-
CC      AD54071). The coding sequences are useful in the gene therapy of
CC      diseases caused by abnormalities of the proteins, e.g. cancer,
CC      inflammatory diseases, osteoporosis or neurological disease.
XX
SQ      Sequence 2208 BP; 660 A; 454 C; 451 G; 643 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      3.3e-126      Length:      2208
Score:      1204.00      Matches:      239
Percent Similarity:      89.51%      Conservative:      0
Best Local Similarity:      89.51%      Mismatches:      4
Query Match:      86.93%      Indels:      24
DB:      10      Gaps:      1

US-10-649-273-2_COPY_148_414 (1-267) x AD52832 (1-2208)
QY      1 MetGluAlaHisAlaIleuThrIleArGLeuThrAsnIyValGluPheProPheLeuVal 20
DB      785 ATGGAGGCGTCAGACCTTATCTATAGGTTGACCAAAATGTAATTTCTTTTATAGTT 844
QY      21 LeuIleuIleSerGIyGIyHisCysLeuLeuAlaIleuValGIyValSerAspPheIy 40
DB      845 CTTTGATTTCTGGAGGTCATCTGTCTGTGGCATTAAGTTCAAGAGATTTCAAGATTTCTG 904
QY      41 LeuIleuGIyISerIleuAspIleAlaProGIyAspMetLeuAspIyValAlaArGAr 60
DB      905 CTTTGGAAAGTCTTTGGACATAGACACAGGATGATCATCTTGACAAAGGTGGCAAGAGA 964
QY      61 LeuSerIleIleIyHisIProGIyCysSerThrMetSerGIyGIyIyValIleGIyHis 80
DB      965 CTTCCTTAATAATAACATCCACAGTGCCTCACCATAGTGGTGGAAACCATAGAAAT 1024
QY      81 LeuAlaIySGInGIyAaNaArGpHeHiePheAspIleIySProProLeuHiHiaIaIyS 100
DB      1025 TTGGCCCAAGAGAGAAATAGATTTCATTGACATCAAACTCCCTTCATCATCTCTAAA 1084
QY      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIySileIleMetIySlyS 120
DB      1085 AATTGTGATTTTCTTTTACTGAGCTTCACACGTTACTGATTAATAATTAATGAAGAAAG 1144
QY      121 GluIySGInGIyGIleGluIySGIyGlnIleuSerSerAlaIaAspIleAlaIa 140
DB      1145 GAAAAAGAGAGAGATTGAGAGAGGCAAAATCTGTCTTACGACGACACATTCGTGCC 1204
QY      141 ThrValGlnHisThrMetAlaCysHisIleuValIySArGThrHisArGAlaIleuPhe 160
DB      1205 ACAGTACAGACACAAATGGCATGTCACTTGTGAAAAGAACACATCGGGCTATTCTGTTT 1264
QY      161 CysIySGInArGAspLeuPheProGlnAsnAsnAlaValIleuValAlaSerGIyGIyVal 180
DB      1265 TGTAAAGCAGAGAGACTGTGTAACCTCAAAATATGCACTGTGTCATCTGTGTGTC 1324
QY      181 AlaSerAsnPheTyrlleArGArGAlaIleuGlnIleuThrAsnAlaThrGlnCysThr 200
DB      1325 GCAAGTAACTTCTATATCCGACAGCTCGAATAATTTTAAACAACGCAACACAGTGCAC 1384
QY      201 LeuIleuCysProProArGLeuCysThrAspAsnGIyIleMetIleAlaTrpAsnGIy 220
DB      1385 TTGTGTGTCTCTCTCCACAGCTATGCACTGATTAATGGAATTAATGATTCAC----- 1435
QY      221 IleGluArGLeuArGAlaGIyLeuGIyIleuHisAspIleGIyIleArGTrGIu 240
```



Db 1435 ----- 1435  
Qy 241 ProlvsCysProlenGlyValAsp1IleSerlySGlyValGlyGluAlaSer1IleLysVal 260  
Db 1436 ---TGAGTCTCTTGGAGTACATATCAAAAGAGTTGGAGAGCTTCCATATAAGTA 1492  
Qy 261 ProGlnLeuLysMetGluIle 267  
Db 1493 CCACAATTAAAAATGAGATA 1513  
RESULT 10  
ADQ24627  
ID ADQ24627 standard; DNA; 2890 BP.  
XX  
AC ADQ24627;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7447.  
XX  
KM soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
KW de.  
XX  
OS Homo sapiens.  
XX  
PN MO2004048938-A2.  
XX  
PD 10-JUN-2004.  
XX  
PF 26-NOV-2003; 2003WO-US038193.  
XX  
PR 26-NOV-2002; 2002US-0429739P.  
XX  
PA (PROT-) PROTEIN DESIGN LABS INC.  
XX  
PI Aziz N, Ginsburg WM, Zlotnick A;  
XX  
DR WPI; 2004-441208/41.  
XX  
PT Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX  
PS Example 2; SEQ ID NO 7447; 210bp; English.  
XX  
XX The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX  
SQ Sequence 2890 BP; 869 A; 609 C; 611 G; 789 T; 0 U; 12 Other;  
XX  
Alignment Scores:  
Pred. No.: 4,83e-126 Length: 2890  
Score: 1204.00 Matches: 239  
Percent Similarity: 89.51% Conservative: 0  
Best Local Similarity: 89.51% Mismatches: 4  
Query Match: 86.93% Indels: 24  
DB: 12 Gaps: 1  
US-10-649-273-2\_COPY\_148\_414 (1-267) x ADQ24627 (1-2890)  
Qy 1 MetGluAlaHisAlaLeuThr1IleArgLeuThrAlaLysValGluPheProPheLeuVal 20

Db 1442 ATGGAGGCTCAGCAGCTTACTATAGTTGACCAATTAAGTAATTCCTTTTATAGTT 1501  
Qy 21 LeuLeu1IleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
Db 1502 CTTTGAATTTCTGGAGGTACAGTGTCTGTTGGCATTAAGTTCAAGAGATTTCAAGTTTCTCG 1561  
Qy 41 LeuLeuGlyLysSerLeuAsp1IleAlaProGlyAspMetLeuAspLysValAlaArgAla 60  
Db 1562 CTTCTTGGAAAGTCTTTGGACATATGACACAGAGTGCATCTTGCACAAAGGTGGCAAGACA 1621  
Qy 61 LeuSerLeu1IleLysHisProGlyLysSerThrMetSerGlyGlyLysAla1IleGlnHis 80  
Db 1622 CTTCTTAAATTAACATCCAGATGCTCCACCATGATGAGTGGGAAACCATAGAACAT 1681  
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisPheAsp1IleLysProPheLeuHisHisAlaLys 100  
Db 1682 TTGGCCAAACAAAGAAATGATTTCACTTTTGACATCAACCTCCCTTGACATATGCTAA 1741  
Qy 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLys1IleIleMetLysLys 120  
Db 1742 AATTGGAATTTCTTTTACTGGACTTCAACAGCTTACTGATTAATAATTAATGAAG 1801  
Qy 121 GlyLysGluGluGly1IleGlyLysGln1IleLeuSerSerAlaAlaAsp1IleAla 140  
Db 1802 GAAAAAGAGAAAGTAATGAGAAAGGGCAAAATCTGTCTTCAGCACAGCATTTGCTGCC 1861  
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAla1IleLeuPhe 160  
Db 1862 ACAGTACGACACAAATGAGCATGTCATCTTGGAAGAAACACATGGGCTATTCGT 1921  
Qy 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
Db 1922 TGTAAAGCAGAGAGACTGTGTTACTCAATAATAGTACGTGTTGCATCTGTGGTGTCTC 1981  
Qy 181 AlaSerAsnPheTyrlIleArgAlaLeuGlu1IleLeuThrAsnAlaThrGlnCysThr 200  
Db 1982 GCAGTAATCTTGATCCGACAGAGCTCTGAAATTTTAACAAAACCAACACAGTCACT 2041  
Qy 201 LeuLeuCysProProProArgLeuCysThrAspAsnGly1IleMet1IleAlaTrpAsnGly 220  
Db 2042 TTGTTGTGTCCTCTCCACGACTATGACTGTAATAGGCAATTAATGATGCA----- 2092  
Qy 221 IleGluArgLeuArgAlaGlyLeuGly1IleLeuHisAsp1IleGluGly1IleArgTyGlu 240  
Db 2092 ----- 2092  
Qy 241 ProlvsCysProlenGlyValAsp1IleSerlySGlyValGlyGluAlaSer1IleLysVal 260  
Db 2093 ---TGATGTCTCTTGGAGTACATATCAAAAGAGTTGGAGAGCTTCCATATAAGTA 2149  
Qy 261 ProGlnLeuLysMetGluIle 267  
Db 2150 CCACAATTAAAAATGAGATA 2170  
RESULT 11  
ADE31345/c  
ID ADE31345 standard; DNA; 3358 BP.  
XX  
AC ADE31345;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human diagnostic and therapeutic polynucleotide (dthp), SEQ ID NO 100.  
XX  
XX diagnostic and therapeutic polynucleotide; dthp; antiarteriosclerotic;  
XX antiinflammatory; cerebroprotective; antidiabetic; antidiabetic;  
XX immunosuppressive; neuroprotective; nootropic; neuroleptic; tranquilizer;  
XX osteopathic; antirheumatic; antirheumatic; cytostatic; hepatotropic;  
XX vitamin; haemostatic; anti-HIV; antithyroid; thyromimetic;  
XX dermatological; antibacterial; fungicide; antiparasitic; anticonvulsant;  
XX thrombolytic; anticoagulant; anorectic; vasotropic; antitumor;  
XX gene therapy; protein replacement therapy; human; gene; ds.

OS Homo sapiens.  
 XX MO2003062376-A2.  
 XX  
 XX 31-JUL-2003.  
 XX  
 XX 13-JAN-2003; 2003WO-US001096.  
 XX  
 XX 16-JAN-2002; 2002US-0349384P.  
 XX 17-JAN-2002; 2002US-0349413P.  
 XX 17-JAN-2002; 2002US-0349946P.  
 XX  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JL,  
 XX Yu JY, Thason O, Yap PE, Ameshey SR, Dam TC, Liu TP, Gerslin EH,  
 XX Peralta CH, Lewis SA, Chen A, Marwaha R, Lan RV, Uraehka ME,  
 XX Kristnam SR, Kolluru V, Panesar IS;  
 XX WPI; 2003-636732/60.  
 XX P-PSDB; ADE31156.  
 XX  
 XX New human diagnostic and therapeutic polynucleotides and polypeptides,  
 XX useful for diagnosing, treating or preventing e.g. leukemia, brain  
 XX cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke  
 XX or Alzheimer's.  
 XX  
 XX Claim 1; SEQ ID NO 100; 634bp; English.  
 XX  
 XX The invention relates to a novel isolated human diagnostic and  
 XX therapeutic polynucleotide (designated ditp). The novel ditp  
 XX polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798  
 XX base pairs fully defined in the specification; a polynucleotide  
 XX comprising a naturally occurring polynucleotide sequence at least 90%  
 XX identical to the ditp polynucleotide; a polynucleotide complementary to  
 XX the ditp polynucleotide or its polynucleotide which is at least 90%  
 XX identical; or an RNA equivalent of any of the polynucleotides mentioned  
 XX above. The ditp polynucleotides have the following activities:  
 XX antiarteriosclerotic, antiinflammatory, cerebroprotective, antilipemic,  
 XX antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,  
 XX tranquilizer, osteopathic, antiarthritic, antineumatic, cytostatic,  
 XX hepatotropic, virucide, haemostatic, anti-HIV, antithyroid, chryomimetic,  
 XX dematological, antibacterial, fungicide, antiparasitic, anticonvulsant,  
 XX thrombolytic, anticoagulant, anorectic, vasotropic, and anticancer. The  
 XX novel DITP polynucleotides polypeptide can be used in gene therapy and  
 XX protein replacement therapy. The ditp polynucleotides or DITP  
 XX polypeptides are useful for diagnosing, preventing or treating diseases  
 XX associated with the expression of human molecules. In particular, these  
 XX diseases include cancers (e.g. adenocarcinoma, leukemia, melanoma, brain  
 XX cancer, breast cancer, cervix cancer, bone cancer, liver cancer, lung  
 XX cancer) or other cell proliferative disorders (e.g. arteriosclerosis,  
 XX atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary  
 XX thrombocytopoenia), autoimmune/inflammatory disorders (e.g. AIDS,  
 XX Addison's disease, thyroiditis, Crohn's disease, Graves' disease,  
 XX Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid  
 XX arthritis, scleroderma, systemic lupus erythematosus), infections (e.g.  
 XX viral, bacterial, fungal or parasitic infection), developmental disorders  
 XX (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.  
 XX thrombosis, hypopituitarism, hypogonadism, gigantism, goiter) metabolic  
 XX disorders (e.g. hypercholesterolaemia, hypoglycaemia, diabetes,  
 XX hyperlipidaemia, obesity), neurological disorders (e.g. ischemic  
 XX cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,  
 XX Huntington's disease, Parkinson's disease, Creutzfeldt-Jacob disease,  
 XX anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers),  
 XX transport disorders (e.g. akinesia or multidrug resistance), or  
 XX connective tissue disorders (e.g. Paget's disease or rickets). This  
 XX polynucleotide sequence represents one of the human ditp DNA sequences  
 XX of the invention.  
 XX  
 XX Sequence 3358 BP; 1105 A; 577 C; 601 G; 1075 T; 0 U; 0 Other;  
 XX  
 XX Alignment Scores:

Pred. No.:	2,99e-102	Length:	3358
Score:	995.50	Matches:	250
Percent Similarity:	44.82%	Conservative:	1
Best Local Similarity:	44.64%	Mismatches:	7
Query Match:	71.88%	Indels:	307
DB:	10	Gaps:	2

US-10-649-273-2\_COPY\_148\_414 (1-267) x ADE31345 (1-3358)

QY	10	LeuThrAsnIysValGluPheProPheLeuValLeuLeuIleSerGlyIghIstCysLeu	29
DB	3067	GTACCAATTAAGTATGATTTCTTTTATGTTCT-TTGATTTCTGAGGCTCATGCTG	3009
QY	30	LeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyPheSerLeuAspIleAla	49
DB	3008	TGGGCAATTAGTTCAAGAGAGTTTCAGATTTTCTGCTTCTTGGAAGTC-TTGACATAGCA	2950
QY	50	ProGlyAspMet-----	53
DB	2949	CCAGGTGACATGCGTTGACAGAGTAATTAAGTAATTTCTCATCTTTTGTATATGT	2890
QY	53	-----	53
DB	2889	TGTCCATTCACTAAGTAGCAATAGATGTCTACCACTTCACTTAATTTCTGAA	2830
QY	53	-----	53
DB	2829	TTTTATCTTAGTAAGTGAACAAATTCACATATGTGTAGAGAAAAATAGAAAGCAGTA	2770
QY	53	-----	53
DB	2769	GTACAGCAATTTATTAATTTCTTAGCCTTTCTTAATAAAGTATAGAGTTTCATATCTGTA	2710
QY	53	-----	53
DB	2709	CATAAAGCTGAATATGTTTGCAGATACATATGATTTTGTCCAAATATATATGTGAA	2650
QY	53	-----	53
DB	2649	AGAACTGCTGCTGAATCTAATCTACTGCAAAAAAGTAAATATAGAAATATATATAGA	2590
QY	53	-----	53
DB	2589	TTAACAATAGACATTAAGATGATGCAATGACAGATTAATATACACAAATTAATTACACCA	2530
QY	53	-----	53
DB	2529	CAGACAGGTGCCCCCGCACCCCCCTTTGTTTGAATACTACAGAGGCTACATGATAT	2470
QY	53	-----	53
DB	2469	ATAGAACTACAAACAACAACAGACAGCTGCTCCACAGTGAATATATAGAAATATAGG	2410
QY	53	-----	53
DB	2409	ACAAGTTTATTTATTTAGAGTTTCATCATTAAGCAATTTATTCACATTAAGCCATTTT	2350
QY	53	-----	53
DB	2349	CCAACCAATAGAGCAAAACATAGACAGGGGCAATGATGCTTTATTTGTCGGGTC	2290
QY	53	-----	53
DB	2289	ATCATTAAGAACAGGTTGTCTGCTTACCTGAATATACACTATATAGTCTATATTTGCCAA	2230
QY	53	-----	53
DB	2229	GATATGATGTTTATTTATTCATTCAGGGGTTTTTTTGTGTATGTAATTTCAATTAATTT	2170
QY	54	-----	54
DB	2169	CCTTTCATCTTTTGTGTTTCAGAGTATTAATTTATATGACTTAATAAATATATGTTCTTT	2110

QY	55	AspIysValAlaAaGdGLeuSerLeuIleuYehiSPProGluCySethrMetSerGly	74
Db	2109	GATNG-GTGGCAAGAGACTTTCTTTAATAAACAATCCAGAGTCTCCACCAATGAGTGT	2051
QY	75	-GlyIysAlaIleGluhiSleuAlaIysGInGlyAsnaRghehiSphaepIleYsP	94
Db	2050	GGGAAAGCCATGAAACATTTGGCCAAACAAGAAATGATTCTCATTTTGAATCAAAAC	1991
QY	94	oProLeuhiShiSaIalySaenCySaRphSerPheThrGlyLeuGlnhiSvalThrAs	114
Db	1990	TCCCTTGATCATGCTAAAAAATGGAATTTTCTTTTACTGGAATTCACAAGTTACGA	1931
QY	114	PlYsIleIleMetIySlySGluYSGluGlu-----	124
Db	1930	TAAATATATATGAAAAAGAAAAAGAAAGCAAGTATATTCTTAATTAGTAAAGTTGAACA	1871
QY	125	-----GlyIleGluYsG	129
Db	1870	GATAAATATTCCTGGATTGTGCTTAAAAATAGCTGTGCTATTTCGACAGTATTGGAAGG	1811
QY	129	IyGInIleLeuSerSerAlaIaAspIleAlaIaIaThrValGlnhiSthrMetAlaCySh	149
Db	1810	GGCAAAATCTCTCTTCACACAGACATTTGCTGCCACAGTACAGACACAATGGCATGTC	1751
QY	149	ISleuValIySaRgThrhiSaRgAlaIleIleuPheCyAlaRgSaRgAspLeuPProG	169
Db	1750	ATCTTGTAAGAAAGAACACATCGGGCTATTCTGTTTGTAAAGCAAGAGACTGTAACTC	1691
QY	169	IInaSnAsnaIaValIleuValAlaSerGlyGlyValAlaSerAsnPhetyrIleArgArGa	189
Db	1690	AAATTAATGACGACTGTGTCATCTGT-GGTGTGGCAAGTAACTTCAATTCGCCAAG	1632
QY	189	IaIeuGluIleIeuThraSnAlaThrGInCySthrLeuLeuCySProPProArgLeuC	209
Db	1631	CTCTGGAATTTTAAAC-AACCAACAACAGTGCACATTTGTGTGTCCTCCACAGATAT	1572
QY	209	ySthraRPaSnGlyIleMetIleAlaIaThraPnaenGlyIleGluArgLeuRgaIaGlyLeuG	229
Db	1572	GCACCTGATTAATGGCATTTATGATGATGATGAATGTATTAAGAAAGATACGTCTGGCTTGG	1513
QY	229	IyIleLeuhiSaSpIleGluGlyIleArgYrGluPProIyCySProLeuGlyValaAPI	249
Db	1512	GCATTTTACATGACATGAAAGGCATCCCTATGAACCAAAATGTCTCTTGAGATGACA	1455
QY	249	IeSerIySGluValGlyGluIaSerIleYsValProGlnLeuIySmetGluIle	267
Db	1452	TATCAAAAGAAAGTTGAGAGAGCTTCATTAAGATCAACAATTAATAATGAGATA	1397
RESULT 12			
ABQ75508	ID	ABQ75508 standard; DNA; 1572 BP.	
XX	AC	ABQ75508;	
XX	DT	07-NOV-2002 (first entry)	
XX	DE	Murine sialoglycoprotease-like gene sequence SEQ ID NO:7.	
XX	KW	Murine; mouse; protease; calcium activated neutral protease type 5;	
XX	KW	CARN5; trypsinase 4; sialoglycoprotease; enzyme; genetic disease;	
XX	KW	neurological; neuropsychological; psychotic illness; transgenic animal;	
XX	KW	gene; ds.	
XX	OS	Mus musculus.	
XX	PN	WO200245491-A2.	
XX	PD	13-JUN-2002.	
XX	PF	05-DEC-2001; 2001WO-US046405.	
XX	PR	06-DEC-2000; 2000US-0251803P.	
XX	PR	06-DEC-2000; 2000US-0251820P.	

PR	13-DEC-2000; 2000US-0255971P.
XX	
PA	(DELT-) DELTAGEN INC.
XX	
PI	Allen KD, Leviten MW;
XX	
DR	WPI; 2002-657389/70.
XX	
PT	Novel transgenic animal, comprising a disruption in protease target gene,
PT	is useful for identifying agents that ameliorates a phenotype associated
PT	with a disruption in a protease target gene.
PS	
PS	Example 3; Fig 7; 62pp; English.
XX	
CC	The present invention describes a non-human transgenic animal (I)
CC	comprising a disruption in a protease target gene (PG) selected from
CC	calcium activated neutral protease type 5 (CAPN5) gene, trypsin 4 gene
CC	and stialoglycoprotease-like gene. Also described is a targeting construct
CC	(II), comprising a first polynucleotide sequence homologous to at least a
CC	first portion of PG, a second polynucleotide sequence homologous to at
CC	least a second portion of PG and a selectable marker. (II) is useful for
CC	producing a transgenic mouse comprising a disruption in a protease target
CC	gene, by introducing (II) into a cell, introducing the cell into a
CC	blastocyst, implanting the resulting blastocyst into a pseudopregnant
CC	mouse, where the pseudopregnant mouse gives birth to a chimeric mouse,
CC	and breeding the chimeric mouse to produce the transgenic mouse. (I) is
CC	useful for identifying an agent that modulates the expression or function
CC	of a protease target gene, by administering an agent to (I) and
CC	determining whether the expression or function of the disrupted protease
CC	target gene in (I) is modulated. (I) is also useful for testing the
CC	efficacy of proposed genetic and pharmacological therapies for human
CC	genetic diseases, such as neurological, neuropsychological or psychotic
CC	illness. The present sequence represents murine stialoglycoprotease-like
CC	gene sequence, which is used in an example from the present invention
XX	
SQ	Sequence 1572 BP; 459 A; 337 C; 340 G; 429 T; 0 U; 7 Other;
Alignment Scores:	
Pred. No.:	2, 67e-101 Length: 1572
Score:	983.00 Matches: 210
Percent Similarity:	84.21% Conservative: 14
Best Local Similarity:	78.95% Mismatches: 38
Query Match:	70.97% Indels: 6
DB:	Gaps: 2
US-10-649-273-2_COPY_148_414 (1-267) x ABQ75508 (1-1572)	
OY	1 MetGuaAHnHAlaleuThrlleArgueThzAsnLysVAlGluPhProPheLeuVal 20       :
Dd	445 ATGAGAGCTCAGCACTGACTATAGGCTCACAAATAAAGTAAACATTTCCTTTTAAAGTT 5038       :
OY	21 LeuLeuLlSeSerClYgLYhIsCYsLeuLeuAlaleuValGInglYValSeraspPheLeu 40       :
Dd	504 CTTTTATTTTCGGGGYGACACTGCCGTGGKRTTAAGTCCAAGGGTTCGCATTTTCCTG 5638       :
OY	41 LeuLeuGLYySerLeuaspLllealAprOGLyasPdeLeuaspLysValAladArg 60       :
Dd	564 CTCCTTGGAAGAAGCTTTGGACAATACACACAGCGCAATCATGTGACAAAGTGGCAAAGAGA 6228       :
OY	61 LeuSerLeuLlleuShIsPrOGLucySerThrmetsargLYgLYValAlleGluHis 80       :
Dd	624 CTTTTCTTTATCAAAATCATCAAGATGTTCTCAAAATGAGTGGTGAAGAGCTATAGAACAG 6838       :
OY	81 LeuAlAlayGlnLYAsnarPpHehlspheAsrLlleusProproLeuHlsHleAlAlays 1008       :
Dd	684 TTGGCAAAAGACGGAAATAGATTCATTTTATCAATCAATCAACCTATAGCAAGAATGCTBAG 7438       :
OY	101 AsnCysaspPheSerPheThrLYleuGlnHlsVALThrAspLysLlellemetLysLys 1208       :
Dd	744 AATTGGCATTTTCTTTCAACGGGAGCTTCAACATTAATCGATTAAGCTAATTAACACACAG 8038       :
OY	121 GluLYGluGluGluLYlleGluYseLYGlnlleLeuSerSerAlaAlaAspLlleAlala 1408       :

Db 804 GAAAGAGAGAGGATGAGAGAGGCAATCTCTCATCATGCTGACAGACATTCCTGCT 863  
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValIysArgThrHisArgAlaIleLeuPhe 160  
Db 864 GCGGTACAGCATGCAACAGCGGCGACCTTGGGAAAGAACATCGCGCTATTCTGTTT 923  
Qy 161 CysIysGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyGlyVal 180  
Db 924 TGCAAGAGAAATTTGCTCTCTCCAGCTAACGAGATTTAGTTGATCTGAGAGTGT 983  
Qy 181 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
Db 984 GCAAGTACTGTATCATCCGAAAGACATTGGAAATTTGTGCAGAAATGCACGATGCACGG 1043  
Qy 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
Db 1044 TTGTTGTCTCCACT-TCAAGACTGTGCATGACATAGGCAATGATTCATGAGATGA 1102  
Qy 221 TleGluArgLeuAlaGlyLeuGlyIleLeu-HisAsp-IleGluGlyIleArgTyrG 240  
Db 1103 ATTGGAAATTAACGCTGCTGCTTGCCTTTACCATGATGATGAGACATTCGGTTATT 1162  
Qy 240 IupProLysCysProLeuGlyVal---AspIleSerLysGlnValGlyGlnAlaSerIleL 259  
Db 1163 AACCCAAATGCTCTTGAAGTGAAGCATTTATCCGAAAGATGGCAGA---AGCTTGC 1219  
Qy 259 ySValProGlnLeu 263  
Db 1220 CCATTAAAGTTA 1233

RESULT 13  
AAS84622  
ID AAS84622 standard; cDNA; 2734 BP.  
XX AAS84622;  
AC  
XX  
XX 13-FEB-2002 (first entry)  
DE  
XX DNA encoding novel human diagnostic protein #20426.  
XX  
XX Human, chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
OS  
XX MO2001.75067-A2.  
PN  
XX 11-OCT-2001.  
PD  
XX 30-MAR-2001; 2001WO-US008631.  
PF  
XX 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYS-E-) HYSSEQ INC.  
XX  
PI Dermanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG20435.  
XX  
XX

New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 1; SEQ ID NO 20426; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed

CC gener. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. Ab564197-AB594564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences  
XX

Sequence 2734 BP; 773 A; 545 C; 639 G; 763 T; 0 U; 14 Other;  
SQ

Alignment Scores:  
Pred. No.: 2, 61e-57 Length: 2734  
Score: 599.00 Matches: 131  
Percent Similarity: 71.57% Conservative: 10  
Best Local Similarity: 66.50% Mismatches: 22  
Query Match: 43.25% Indels: 34  
DB: 5 Gaps: 6

US-10-649-273-2\_COPY\_148\_414 (1-267) x AAS84622 (1-2734)

Qy 71 ThrMetSerGlyGlyValAlaIleGlnHisLeuAlaLysGlnIleAsnArgPheHisPhe 90  
Db 206 ACCCTGCTAGGGGGCCCGCT-----AGATTCAT--- 235  
Qy 91 AspIleLysProProLeuHisHisAlaLys-----AsnCys----- 102  
Db 236 -----CCTATTCTCCGATGAAGATTCAGCGCTGTAGAGAACCTGGCAGCA 283  
Qy 103 -----AepPheSerPheThrGlyLeuGlnHisValThrAsp 114  
Db 284 TCTGCTTCGGGAGGCCAGGAGAGCTTTTACTCATCGCGGAGCAATGGAGATGAG 343  
Qy 115 LysIleIleMetLysLysGlnLysGlnIle-----GlyIleGluLysGlnIleLeu 132  
Db 344 CGCTTACATGAGCAGAGAGCAGCAGCAGAGAGGGGGGATATTGAGAAAGGGCAATCTCG 403  
Qy 133 SerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLys 152  
Db 404 TCTTCAGCAGACATTCGCTGCACAGATACAGCACAATGCAATGTCATCTTGAA 463  
Qy 153 ArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAla 172  
Db 464 AGAACATCGGGCTATTCTGTTTGTGTTTAAAGCAGAGACTGTTACTCAAAATTAATGCA 523  
Qy 173 ValLeuValAlaSerGlyGlyValAlaLysAsnPheTyrIleArgArgAlaLeuGlnIle 192  
Db 524 GTACTGTTGCATCTGTGTGTGTGTGCGAAGTACTTATATCGCAGAGCTCGGAATT 583  
Qy 193 LeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsn 212  
Db 584 TTAACAAAGCAACAGACAGTCACTTGTGTGTGCTCTCCAGCATGCACTGATTAAT 643  
Qy 213 GlyIleMetIleAlaIleTrpAsnGlyIleGlnArgLeuAlaGlyLeuGlyIleLeuHis 232  
Db 644 GGCATTATTAATATGATGATGATGATTAAGATTAAGCTGCGCTTGGCATTTTACAT 703  
Qy 233 AspIleGluGlyIleArgTyrGluProLys-----CysProLeuGlyVal 247  
Db 704 GACATTAAGAGCATCCGCTATTAACCAAAAGTATGTGCTCTTCAGAGGCTG 754

RESULT 14  
ADL86725  
ID ADL86725 standard; DNA; 371 BP.  
XX

AC ADL86725;  
DT 20-MAY-2004 (first entry)  
DE DNA up-regulated in murine multipotent progenitor cells SeqID 3118.  
XX gene potential; multi-lineage; cell commitment; haematopoietic stem cell;  
KM HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;  
KM common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.  
OS Mus sp.  
XX WO2003093445-A2.  
XX  
XX  
PD 13-NOV-2003.  
XX  
PF 05-MAY-2003; 2003WO-US014114.  
XX  
PR 03-MAY-2002; 2002US-0377383P.  
XX  
PA (STOW-) STOWERS INST MEDICAL RES.  
XX  
XX Li L;  
PI  
XX WPI; 2004-022656/02.  
XX  
DR  
PT Classifying an unknown multi-lineage affiliated gene comprises isolating  
PT expressed nucleic acid sequences from the discrete cell sub-populations.  
XX  
XX  
PS Claim 8; SEQ ID NO 3118; 123pp; English.  
XX  
XX This invention relates to a novel method for predicting gene potential by  
CC associating nucleic acid sequences of unknown function with particular  
CC sub-population profiles. Specifically, it refers to classifying an  
CC unknown multi-lineage affiliated gene by collecting hybridisation data  
CC develop a gene expression map, in order to determine the discrete sub-  
CC population where it is expressed. The present invention describes methods  
CC for predicting the lineage commitment of genes associated with the self-  
CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-  
CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors  
CC (CLPs) and common myeloid progenitors (CMPs), which are collectively  
CC referred to as bone marrow stem cells populations. As such, these methods  
CC can be used to identify associated multi-lineage affiliated genes and  
CC hence the underlying molecular mechanisms in physiological haematopoietic  
CC development. This polynucleotide sequence is DNA associated with a murine  
CC MPP sub population of cells of the invention.  
XX  
XX Sequence 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other;

Pred. No.:	1.18e-43	Length:	371
Score:	468.00	Matches:	92
Percent Similarity:	78.86%	Conservative:	5
Best Local Similarity:	74.80%	Mismatches:	26
Query Match:	33.79%	Indels:	0
DB:	12	Gaps:	0
US-10-649-273-2_COPY_148_414 (1-267) x ADL6725 (1-371)			
OY	137	AspIleAlaAlaThrValGlnHisThrmcAlaCysH;slenValuAllysArGThrHisArg	156
Db	2	GACATGCTGCTGCGCGGTACAGCATGCAACAGCTGCCACTTCCGAAAAGAACACATCCG	61
OY	157	AlaIleLeuPheCysAlsGlnArgspLeuPrcGlnAsnAsnAlaValIleuValAla	176
Db	62	GCTATTCTGTTTGGACGAGGAATTTGCTCTCCAGCTAACCGACGATTAAGTGTGA	121
OY	177	SerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaIleuGlnIleuThrAsnAla	196
Db	122	TCTGAGAGTGTGGCAAGTAACCTTGTACATCCGAAAAGCATGGAAATTGTCCGAAAAGCA	161
OY	197	ThrGlnCysThrLeuLeuCysAspProPrcArgLeuCysThrAspAngIylIleMetIle	216

Db	182	ACGAGATGCACGGTTGTCTGCCACTCCAGACACTGTGCATCAGCAATGGCATCAAGAATT	246
Oy	217	AlAtTtPangngYllleGuLxrgLeuhArgAlagIyleUgYllleUhnIsaspIlleglUglY	236
Db	242	GCAATGGAAATGGAATTTGAANNGAATTAAGTCSCNNGCTTGNCNTNTTACNTGANNATAGAAAGC	301
Oy	237	IleaRgrYrGUPLroLYeCyEProLeugLyVaIAepIIseSerlyeSguValGIgYuaLa	256
Db	302	ATCGNNTATGNACCAAAATNTCNTMTTGGAGTAGACATNTCCAGAGAACTTGAGAGAGCT	361
Oy	257	SerIleIys	259
Db	362	GCCATAAAA	370
<b>RESULT 15</b>			
ID	ADL6726	standard; DNA; 371 BP.	
XX	ADL6726;		
AC	ADL6726;		
XX	ADL6726;		
DT	20-MAY-2004	(first entry)	
XX			
DE	DNA up-regulated in murine multipotent progenitor cells SeqID 3119.		
XX			
KW	gene potential; multi-lineage; cell commitment; haematopoietic stem cell;		
KM	HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;		
KW	common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; da.		
XX	Mus sp.		
OS			
PX	WO2003093445-A2.		
PN			
XX	13-NOV-2003.		
PD			
PF	05-MAY-2003; 2003WO-US014114.		
XX			
PR	03-MAY-2002; 2002US-0377383P.		
PA	(STOW-) STOWERS INST MEDICAL RES.		
XX			
PI	L1 L;		
XX			
DR	WPI; 2004-022656/02.		
XX			
PS	Claim 8; SEQ ID NO 3119; 123pp; English.		
XX			
CC	This invention relates to a novel method for predicting gene potential by		
CC	associating nucleic acid sequences of unknown function with particular		
CC	sub-population profiles. Specifically, it refers to classifying an		
CC	unknown multi-lineage affiliated gene by collecting hybridisation data to		
CC	develop a gene expression map, in order to determine the discrete sub-		
CC	population where it is expressed. The present invention describes methods		
CC	for predicting the lineage commitment of genes associated with the self-		
CC	renewing haematopoietic (blood) stem cells (HSCs), as well as the non-		
CC	self renewing multipotent progenitors (MPPs), common lymphoid progenitor		
CC	(CLPs) and common myeloid progenitors (CMPs), which are collectively		
CC	referred to as bone marrow stem cells populations. As such, these methods		
CC	can be used to identify associated multi-lineage affiliated genes and		
CC	hence the underlying molecular mechanisms in physiological haematopoietic		
CC	development. This polynucleotide sequence is DNA associated with a murine		
CC	MPP sub population of cells of the invention.		
XX			
XX			
SQ	Sequence 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other:		
<b>Alignment Scores:</b>			
Pred. No.:	1,18e-43	Length:	371
Score:	468.00	Matches:	92
Percent Similarity:	78.86%	Conservative:	5
Best Local Similarity:	74.80%	Mismatches:	26
Query Match:	33.79%	Indels:	0

DB: 12 Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x ADL86726 (1-371)  
QY 137 AsplleAlaATrValAlNHsThrMetAlaCysHlsleuValIysaArgThrHisArg 156  
Db 2 GACATTGCTGCTGGCGGTACAGCATCAACAGCGTGCACCTTGCGAANAAGAACACATCGC 61  
QY 157 AlalIleuPheCysLysGlnArgAspleuLeuProGlnAsnAsnAlaValIeuValAla 176  
Db 62 GCTATTCTGTTTGGAAAGCAGAAATTGGCTCTCCAGCTAACGCAATATTAGTTGTA 121  
QY 177 SerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaIeuGluIleuThrAsnAla 196  
Db 122 TCTGAGAGGTGTGCAAGTAACCTTGACATCCGAAAGCATTTGGAATTGTCGCAATGCA 181  
QY 197 ThrGlnCysThrIleuLeuCysProProArgIeuCysThrAspAsnGlyIleMetIle 216  
Db 182 ACGCAGTGCACGTTGTGTGTCCACCTCCAGACTGTGCATGACATGCAATGCATGATT 241  
QY 217 AlatPAsnGlyIleGluArgIeuArgAlaGlyLeuGlyIleLeuHisAspIleGly 236  
Db 242 GCATGGAATGGAATTGAANGATTACGTGCNGNCTTNGCCTTNTACNTGANNTAGAAGAC 301  
QY 237 IleArgTyrGluProLysCysProIeuGlyValAspIleSerLysGluValGlyIuAla 256  
Db 302 ATCCNNTATGNACCAAATNTCNTNTTGGAGTAGACATNTCCAGAGAAAGTTGCAGAGCT 361  
QY 257 SerIleLys 259  
Db 362 GCCATATAAA 370

Search completed: November 10, 2005, 17:34:13  
Job time : 611.911 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 10, 2005, 16:52:13 ; Search time 5267.16 Seconds  
(without alignments)  
2456.267 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414  
Perfect score: 1385  
Sequence: 1 MEMHALTRITNKVPEPFLV.....DISKVGESAIKVPOLKMEI 267

Scoring table: BLOSUM62  
Xgapop 10.0 , Ygapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+p2n.model -DEV=x1p  
-Q/cgnt2\_1/USFTO.spool\_p/US10649273/runat.02112005.091338.15550/app.query.fasta.1.1429  
-DB=GenEmbl -QFT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273 @CGN 1.1.7242 @runat.02112005.091338.15550 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WANT TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

GenEmbl:.\*  
1: gb\_ba:.\*  
2: gb\_hlg:.\*  
3: gb\_in:.\*  
4: gb\_om:.\*  
5: gb\_ov:.\*  
6: gb\_pat:.\*  
7: gb\_ph:.\*  
8: gb\_pl:.\*  
9: gb\_pr:.\*  
10: gb\_ro:.\*  
11: gb\_sts:.\*  
12: gb\_sy:.\*  
13: gb\_un:.\*  
14: gb\_vi:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1385	100.0	1416	AR541929	Sequence
2	1385	100.0	1526	AR428809	Sequence
3	1385	100.0	1908	BC011904	Homo sapi
4	1385	100.0	2197	AR428803	Sequence

5	1358.5	98.1	1387	6	AR428808	Sequence
6	1358.5	98.1	1387	9	HS4295148	Homo sapi
7	1348	97.3	1245	6	AX664697	Sequence
8	1348	97.3	1245	6	AX664695	Sequence
9	1231	88.9	1820	6	AX664695	Sequence
10	1234	88.4	1017	10	BC058172	Mus muscu
11	1204	86.9	2208	6	AX713716	Mus muscu
12	1204	86.9	2208	9	AK055441	Mus muscu
13	1116	68.6	1546	10	BC078974	Rattus no
14	950.5	68.6	1109	10	BC051211	Mus muscu
15	939	67.7	1522	5	BX934991	Gallus ga
16	928	67.7	860	5	BX930694	Gallus ga
17	924	67.4	1558	5	BX930693	Gallus ga
18	831.5	60.0	84115	9	AC013468	Homo sapi
19	824.5	59.5	14364	9	AR428807	Sequence
20	666	48.1	249601	2	AC114153	Rattus no
21	666	48.1	308652	2	AC121478	Rattus no
22	652.5	47.1	256751	10	AC122825	Mus muscu
23	610	44.0	875	6	CQ721898	Sequence
24	488.5	35.3	1474	3	AK113378	Clona int
25	477.5	34.5	121251	5	AL591593	Zebrafish
26	477.5	31.4	117322	5	AL672217	Zebrafish
27	415.5	30.0	1576	3	AY051882	Drosophila
28	415.5	30.0	1601	6	CQ606432	Sequence
29	415.5	30.0	3656	6	CQ606431	Sequence
30	415.5	30.0	14679	2	AC018262	Drosophila
31	415.5	30.0	180263	3	AC010671	Drosophila
32	415.5	30.0	207432	3	AE003513	Drosophila
33	408	29.5	1672	8	AK070912	Oryza sat
34	407.5	29.4	1443	8	AY024338	Arabidops
35	407.5	29.4	1474	8	AY117283	Arabidops
36	407.5	29.4	1567	8	AY063864	Arabidops
37	401.5	29.0	1557	8	AY084577	Arabidops
38	389	28.1	110000	2	AP006501	Continuation (9 of
39	369	26.6	260271	1	AE017258	Molbachia
40	356.5	25.7	333800	1	SM559192	Mus muscu
41	356	25.7	20021	10	AF367967	Mus muscu
42	356	25.7	179252	10	AF131205	Mus muscu
43	352	25.4	349116	1	AP003003	Mesophilzo
44	349	25.2	110000	1	AE017197	Continuation (2 of
45	342	24.7	756	6	AX685201	Sequence

## ALIGNMENTS

RESULT 1	AR541929	1416 bp	linear	PAT 08-OCT-2004
LOCUS	AR541929	Sequence 177 from patent US 6743619.		
DEFINITION	AR541929			
ACCESSION	AR541929.1	GI:53934009		
VERSION	AR541929.1			
KEYWORDS	Unknown.			
SOURCE	Unknown.			
ORGANISM	Unclassified.			
REFERENCE	1 (bases 1 to 1416)			
AUTHORS	Tang,Y.T., Zhou,P., Goodrich,R., Liu,C., Asundi,V., Ren,F., Zhang,Y., Zhao,Q.A., Yang,Y., Xue,A.D., Wehrman,T., Wang,J.-R., Wang,D. and Drmanac,R.T.			
TITLE	Nucleic acids and polypeptides			
JOURNAL	Patent: US 6743619-A 177 01-JUN-2004;			
FEATURES	Location/Qualifiers			
source	1..1416			
ORIGIN	/organism="unknown"			
	/mol_type="genomic DNA"			

Alignment Scores:  
Pred. No.: 1.37e-126 Length: 1416  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservatve: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0



DB: 6 Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x AR541929 (1-1416)  
QY 1 MetGluAlaHisAlaLeuThr11leargLeuThrAsnLysVal1GluPheProPheLeuVal 20  
Db 502 ATGGAGGCTCAATGCACTTACTATTAGGTGACCAATTAAGTAAGTAATTCCTTTTACTT 561  
QY 21 LeuLeu11SerGlyGlyHisCysLeuLeuAlaLeuVal1Gln1Val1SerAspPheLeu 40  
Db 562 CTTTGTGATTTTGGAGGTCACCTGCTGTGGCATTAGTTCAGAGAGTTTCAGATTTTCTG 621  
QY 41 LeuLeuGlyLysSerLeuAsp11leAlaProGlyAspMetLeuAspLysVal1AlaArgArg 60  
Db 622 CTTCTTGAAAGTCTTTGGACATAGCACCAAGTGACATGCTTGACCAAGGTGGCAAGAGA 681  
QY 61 LeuSerLeu11elyHisProGlyLysSerThrMetSerGlyGlyVal11leGluHis 80  
Db 682 CTTTCTTTAATAAATCAATCCAGAGTCTCCACCATGAGTGGGAAAGCCATAGAACAT 741  
QY 81 LeuAlaLysGln1GlyAsnArgPheHisPheAsp11elysProProLeuHisHisAlaLys 100  
Db 742 TTGGCCAAACAGAAATAGATTTCATTTTGACATCAAACTCCCTTGCAATGCTTAA 801  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisVal1ThrAspLys11leMetLysLys 120  
Db 802 AATTGTGATTTTCTTTTACTGAGCTTCAACCGTTACTGATTAATAATATATGAAAAAG 861  
QY 121 GluLysGln1Gly11leGluLysGlyGln11leLeuSerSerAlaAlaAsp11leAla 140  
Db 862 GAAAAAGAGGAAGGATTTGAGAAAGGGCAAAATCCTGCTTCAGACAGACATTCCTGCC 921  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAla11leLeuPhe 160  
Db 922 ACAGTACAGCACACATGCAATGTCATCTTGAAAAAACAACATCGGCTATTCGTTT 981  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal1leuVal1AlaSerGlyGlyVal 180  
Db 982 TGTAAAGCAGAGACTTTTACTTCAAAATATATGACATGCTGTTGCATCTGCTGCTGC 1041  
QY 181 AlaSerAsnPheTyrl1eArgArgAlaLeuGln11leLeuThrAsnAlaThrGlnCysThr 200  
Db 1042 GCAAGTAACTTCTATATCCGAGAGCTCTGGAAATTTTAAACAACGCAACACAGTGCAC 1101  
QY 201 LeuLeuCySPProProPheArgLeuCySPThrAspAsnGly11leMet11leAla1TrpAsnGly 220  
Db 1102 TTGTGTGTCTCTCCCTCCAGACTATGCACTGATTAATGCAATTAATGCAATGCAATGCT 1161  
QY 221 11leGluArgLeuArgAlaGlyLeuGly11leLeuHisAsp11leGluGly11leArgTyrglu 240  
Db 1162 ATTGAAGACTACGTCGCTGGCTTGGCAATTTTACATGACATGAAAGCATCCGCTATGAA 1221  
QY 241 ProLysCySPProLeuGlyVal1Asp11leSerLysGluVal1GlyGluAlaSer11elysVal 260  
Db 1222 CCAAAATGCTCCCTTGAGTAGACATATCAAAAGAAAGTTGGAGAAAGCTTCCATTAAGTA 1281  
QY 261 ProGlnLeuLysMetGlu11le 267  
Db 1282 CCACAATTAATAATGAGATA 1302  
RESULT 2  
AR428809 1526 bp DNA linear PAT 18-DEC-2003  
LOCUS AR428809 Sequence 23 from patent US 6642041.  
ACCESSION AR428809  
VERSION AR428809.1 GI:40188595  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1526)  
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.  
TITLE Polynucleotides encoding a novel metalloprotease, MP-1

JOURNAL Patent: US 6642041-A 23 04-NOV-2003;  
FEATURES Location/Qualifiers  
source 1..1526  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Alignment Scores:  
Pred. No.: 1..5e-126 Length: 1526  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x AR428809 (1-1526)  
QY 1 MetGluAlaHisAlaLeuThr11leargLeuThrAsnLysVal1GluPheProPheLeuVal 20  
Db 1 ATGGAGGCTCAATGCACTTACTATTAGGTGACCAATTAAGTAAGTAATTCCTTTTACTT 60  
QY 21 LeuLeu11SerGlyGlyHisCysLeuLeuAlaLeuVal1Gln1Val1SerAspPheLeu 40  
Db 61 CTTTGTGATTTTGGAGGTCACCTGCTGTGGCATTAGTTCAGAGAGTTTCAGATTTTCTG 120  
QY 41 LeuLeuGlyLysSerLeuAsp11leAlaProGlyAspMetLeuAspLysVal1AlaArgArg 60  
Db 121 CTTCTTGAAAGTCTTTGGACATAGCACCAAGTGACATGCTTGACCAAGGTGGCAAGAGA 180  
QY 61 LeuSerLeu11elyHisProGlyLysSerThrMetSerGlyGlyVal11leGluHis 80  
Db 181 CTTTCTTTAATAAATCAATCCAGAGTCTCCACCATGAGTGGGAAAGCCATAGAACAT 240  
QY 81 LeuAlaLysGln1GlyAsnArgPheHisPheAsp11elysProProLeuHisHisAlaLys 100  
Db 241 TTGGCCAAACAGAAATAGATTTCATTTTGACATCAAACTCCCTTGCAATGCTTAA 300  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisVal1ThrAspLys11leMetLysLys 120  
Db 301 AATTGTGATTTTCTTTTACTGAGCTTCAACCGTTACTGATTAATAATATATGAAAAAG 360  
QY 121 GluLysGln1Gly11leGluLysGlyGln11leLeuSerSerAlaAlaAsp11leAla 140  
Db 361 GAAAAAGAGGAAGGATTTGAGAAAGGGCAAAATCCTGCTTCAGACAGACATTCCTGCC 420  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAla11leLeuPhe 160  
Db 421 ACAGTACAGCACACATGCAATGTCATCTTGAAAAAACAACATCGGCTATTCGTTT 480  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal1leuVal1AlaSerGlyGlyVal 180  
Db 481 TGTAAAGCAGAGACTTTTACTTCAAAATATATGACATGCTGTTGCATCTGCTGCTGC 540  
QY 181 AlaSerAsnPheTyrl1eArgArgAlaLeuGln11leLeuThrAsnAlaThrGlnCysThr 200  
Db 541 GCAAGTAACTTCTATATCCGAGAGCTCTGGAAATTTTAAACAACGCAACACAGTGCAC 600  
QY 201 LeuLeuCySPProProPheArgLeuCySPThrAspAsnGly11leMet11leAla1TrpAsnGly 220  
Db 601 TTGTGTGTCTCTCCCTCCAGACTATGCACTGATTAATGCAATTAATGCAATGCAATGCT 660  
QY 221 11leGluArgLeuArgAlaGlyLeuGly11leLeuHisAsp11leGluGly11leArgTyrglu 240  
Db 661 ATTGAAGACTACGTCGCTGGCTTGGCAATTTTACATGACATGAAAGCATCCGCTATGAA 720  
QY 241 ProLysCySPProLeuGlyVal1Asp11leSerLysGluVal1GlyGluAlaSer11elysVal 260  
Db 721 CCAAAATGCTCCCTTGAGTAGACATATCAAAAGAAAGTTGGAGAAAGCTTCCATTAAGTA 780  
QY 261 ProGlnLeuLysMetGlu11le 267  
Db 781 CCACAATTAATAATGAGATA 801

RESULT 3  
BC011904  
LOCUS  
DEFINITION  
Homo sapiens O-6-acylglycoprotein endopeptidase-like 1, mRNA (CDNA  
clone MGC:20293 IMAGE:4121450), complete cds.  
ACCESSION  
BC011904  
VERSION  
BC011904.2 GI:40225818  
KEYWORDS  
MGC.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 1908)  
Strausberg,R.,L., Felgould,E.A., Grouse,L.H., Derge,J.G.,  
Klausner,R.D., Collins,F.S., Wagner,L., Shennan,C.M., Schuler,G.D.,  
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heide,F.,  
Datchenko,L., Marsina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
Schaefer,T.E., Brownstein,M.J., Usdin,T.B., Toshitsuki,S.,  
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,  
Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,  
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
Morley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,  
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,  
Sanchez,A., Whitling,M., Madan,A., Young,A.C., Shevchenko,Y.,  
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalls,D.E.,  
Schmerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 1908)  
Strausberg,R.  
Direct Submission  
Submitted (30-JUL-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
On Dec 19, 2003 this sequence version replaced gi:15080281.  
Contact: MGC help desk  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@hgt.nih.gov](mailto:nisc_mgc@hgt.nih.gov)  
Akter,N., Ayala,K., Beckstrom-Sternberg,S.M., Benjamin,B.,  
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooke,S.,  
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,  
Hansen,N., Ho,S.-L., Karins,E., Kwong,P., Latic,J., Legaspi,R.,  
Maduro,Q.L., Mastello,C., Masker,B., Mastrian,S.D., McCloskey,J.C.,  
McDowell,J., Pearson,R., Stantirlop,S., Thomas,P.J., Touchman,J.W.,  
Teurigne,C., Vogt,J.L., Walker,M.A., Weherby,K.D., Wiggins,L.,  
Young,A., Zhang,L.-H. and Green,E.D.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LMNC at: <http://image.llnl.gov>  
Series: IRAL Plate: 28 Row: 1 Column: 22.  
Location/Qualifiers  
1..1908  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:20293 IMAGE:4121450"  
/tissue\_type="Muscle, rhabdomyosarcoma"

misc\_feature  
/clone\_11b="NIH\_MGC\_17"  
/lab\_host="DH10B-R"  
/note="vector: pOTB7"  
1..1908  
/gene="OSGEP1"  
/db\_xref="LOCUSID:64172"  
104..1348  
/gene="OSGEP1"  
/codon\_start=1  
/product="OSGEP1 protein"  
/protein\_id="AAH1904.1"  
/db\_xref="GI:15080282"  
/db\_xref="LOCUSID:64172"  
/translation="MLITKTKVGFPEKRRKYEFLEPSFNHPGTLFLAKIVLGLET  
SCDDPAAVVDEGTGVAISQTEVSEHLTKGTIVPAQAQILRENIQRIQVQALASAS  
VSPPLDSIAITLIGALSLGVGSLFSLVGLGKPEFPIPHHEAVALITLTK  
VSPPLVLLISGHCILLVQVSPPLIGKSLIDIRGMDLVKVARSLIIGHPEST  
MSGKATIEHLAKQNRHFDIPPLHNAKCPSPFLGHWTDKTIKKEKEGIEKG  
QILSSAADIATVQHTMACHLVKRTHRALIFPKQRDLRQNNNAVASGVASNFYIR  
RALRLTNATQCTLCPPRLCTDNGIMTAMNGIERLRAGLILHDIIEGIRVEPKCPL  
GVDSKEVGEASIKVPLKMEI"  
215..1111  
/gene="OSGEP1"  
/note="peptidase\_M22, Region: Glycoprotease family"  
/db\_xref="CDD:pfam00814"  
ALIGNMENT Scores:  
Pred. No.: 1,976-126 Length: 1908  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x BC011904 (1-1908)  
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
DB 545 ATGGAGGCTCATGCACTTACTTTAGGCTTGACCAATTAAGTAAATTTCTTTTATGTT 604  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 605 CTTTGAATTTTGAGAGTCACTGCTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCTG 664  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgAG 60  
DB 665 CTTTGGAAAGTCTTTGGACATGACACAGGTGACATGTCGACAAAGGTGCAGAGMAA 724  
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGluHis 80  
DB 725 CTTTCTTTAATTAACATCCAGAGTGTCTCCACATGAGTGTGGAGAAACCATAGACAT 784  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisAspAspIleLysProProLeuHisHisAlaLys 100  
DB 785 TTGGCAAAACAGAAATAGATTTTATTTGACATCAAACTCCCTTGATCATGCTAA 844  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
DB 845 AATTGTGATTTTCTTTTACTGAGACTTCACACGTTACTGATTAATAATATATGAAAAAG 904  
QY 121 GluLysGlnGluGluGlyIleGluLysGlyGlnIleLysSerSerAlaAlaAspIleAlaAla 140  
DB 905 GAAAAAGAGAAAGATTTAGAAAGGGCAAAATCTGTCTTACGACGACAGACATTCCTGCC 964  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
DB 965 ACAGTACGACACACATGCGATGTATCTTGTGAAGAAACACATCGGCTATTCTGTTT 1024  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleValAlaSerGlyGlyVal 180  
DB 1025 TGTAAAGCAGAGAGACTTGTACCTCAAAATATATGACGACTGCTTGTCATCTGCTGTCTC 1084

QY 181 AlAseraSpheTyrlleargArgAlaleuGlulileuThraSpnlaIaThrcInCysThr 200  
Db 1085 GCAAGTAACCTTATATCCGAGAGCTCTGGAAAATTTAAACAAGCAACAGTGCACT 1144  
QY 201 LeuLeuCySPROProPProAArgLeuCyThrAspAsnGlylIleMerllleAlaIaThraSpnly 220  
Db 1145 TTGTGTGTCTCTCCCAAGCACTATGCACTGATTAATGCAATTATGATTCAGATGATGAT 1204  
QY 221 lIleGluArgLeuAArgAlaGlyLeuGlylIleuHIsAspIleGluGlylIleArgTyrglu 240  
Db 1205 ATTGAAGAAGCTACGTCTGCTGGCATTTTACATGACATGAAGGACATCCGCTATGA 1264  
QY 241 ProLysCySPROleuGlyValAspIleSerlySGluValGlyGluAlaSerlleLysVal 260  
Db 1265 CCAAAATGTCCTCTTGAGTAGACATATCAAAAGAAGTTGGAGAAGCTTCCATAAAGTA 1324  
QY 261 ProGlnLeuLysMetGluIle 267  
Db 1325 CCACAATTAATAAATGAGATA 1345

RESULT 4  
AR428803  
LOCUS AR428803 2197 bp DNA linear .PAT 18-DEC-2003  
DEFINITION Sequence 1 from patent US 6642041.  
ACCESSION AR428803  
VERSION AR428803.1 GI:40188589  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2197)  
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.  
TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 6642041-A 1 04-NOV-2003;  
FEATURES  
source 1..2197  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Alignment Scores:  
Pred. No.: 2,33e-126 Length: 2197  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x AR428803 (1-2197)

QY 1 MetGluAlaHIsAlaLeuThrlleArgLeuThraSpnlyValGluPheProPheLeuVal 20  
Db 672 ATGAGAGGCTCATGCACTTACTATTAGGTGACCAATAAAGTAATTCCTTTTAAAGTT 731  
QY 21 LeuLeuIleSerGlyGlyHIsCysLeuLeuAlaLeuValGlnGlyValAspAspPheLeu 40  
Db 732 CTTTGTATTTCTGAGGCTACGTCTGTGTGCAATTAGTTCAAGAGTTTCAATTTTCTG 791  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaIaProGlyAspMetLeuAspLysValAlaArgArg 60  
Db 792 CTTCTTGAAAGCTTTTGACATATGACCAAGTGACATGCTTGACAGAGCTGGCAAGAA 851  
QY 61 LeuSerLeuIleLysHIsProGluCysSerThrMetSerGlyLysAlaIleGluHIs 80  
Db 852 CTTTCTTAATAAACAATCCAGAGTGCTCCACCAATGAGTGAGGAAACCCATAGAACAT 911  
QY 81 LeuAlaLysGlnGlyAsnArgPheHIsPheAspIleLysProProLeuHIsHIsAlaLys 100  
Db 912 TTGGCAAAACAAGAAATTAATTCATTTTGACATCAAACTCCCTTGACATCATGTATA 971  
QY 101 AsnCyAspPheSerPheThrGlyLeuGlnHIsValIThrAspLysIleIleMetLysLys 120  
Db 972 AATTGTGATTTTCTTTTACTGCACTTCAACAGCTTACTGATTAATAATATGAAAAAG 1031

QY 121 GluLysGluGluGlylIleGluLysGlyGlnIleLeuSerSerAlaIaAspIleAlaIa 140  
Db 1032 GAAAAAGAGAAAGTATTGAGAAAGGGCAAACTGTCTTCAGACAGACAACTTGTCTGCC 1091  
QY 141 ThrValGlnHIsThrMetAlaCysHIsLeuValLysArgThrHIsArgAlaIaIleLeuPhe 160  
Db 1092 ACAGTACAGCACACAATGGCAGTCACTTGTGAAGAACAATCGGCTATTCGTGT 1151  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAlaValIleuValAlaSerGlyGlyVal 180  
Db 1152 TGTAAACAGAGACCTGTACTTCAAAATTAATGCAATGCTGTTGATCTGCTGTGTC 1211  
QY 181 AlAseraSpheTyrlleargArgAlaleuGluIleLeuThraSpnlaIaThrcInCysThr 200  
Db 1212 GCAAGTAACCTTATATCCGAGAGCTCTGGAAAATTTAAACAAGCAACAGTGCACTG 1271  
QY 201 LeuLeuCySPROProPProAArgLeuCyThrAspAsnGlylIleMerllleAlaIaThraSpnly 220  
Db 1272 TTGTGTGTCTCTCCCAAGCACTATGCACTGATTAATGCAATTATGATTCAGATGATGAT 1331  
QY 221 lIleGluArgLeuAArgAlaGlyLeuGlylIleuHIsAspIleGluGlylIleArgTyrglu 240  
Db 1332 ATTGAAGAAGCTACGTCTGCTGGCATTTTACATGACATGAAGGACATCCGCTATGA 1391  
QY 241 ProLysCySPROleuGlyValAspIleSerlySGluValGlyGluAlaSerlleLysVal 260  
Db 1392 CCAAAATGTCCTCTTGAGTAGACATATCAAAAGAAGTTGGAGAAGCTTCCATAAAGTA 1451  
QY 261 ProGlnLeuLysMetGluIle 267  
Db 1452 CCACAATTAATAAATGAGATA 1472

RESULT 5  
AR428808  
LOCUS AR428808 1387 bp DNA linear .PAT 18-DEC-2003  
DEFINITION Sequence 21 from patent US 6642041.  
ACCESSION AR428808  
VERSION AR428808.1 GI:40188594  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1387)  
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.  
TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 6642041-A 21 04-NOV-2003;  
FEATURES  
source 1..1387  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Alignment Scores:  
Pred. No.: 5.43e-124 Length: 1387  
Score: 1358.50 Matches: 266  
Percent Similarity: 91.10% Conservative: 0  
Best Local Similarity: 91.10% Mismatches: 1  
Query Match: 98.09% Indels: 25  
Gaps: 1

US-10-649-273-2\_COPY\_148\_414 (1-267) x AR428808 (1-1387)

QY 1 MetGluAlaHIsAlaLeuThrlleArgLeuThraSpnlyValGluPheProPheLeuVal 20  
Db 465 ATGAGAGGCTCATGCACTTACTATTAGGTGACCAATAAAGTAATTCCTTTTAAAGTT 524  
QY 21 LeuLeuIleSerGlyGlyHIsCysLeuLeuAlaLeuValGlnGlyValAspAspPheLeu 40  
Db 525 CTTTGTATTTCTGAGGCTACGTCTGTGTGCAATTAGTTCAAGAGTTTCAATTTTCTG 584  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaIaProGlyAspMetLeuAspLysValAlaArgArg 60

```
Db 585 CTTCTGGAAAGTCTTTGGACATAGCACAGGTGACATGCTTGACAGGTGGCAAGACA 644
Qy 61 LeuSerLeuIlelyHisProGluCySerThrMetSerGlyLyalaIleGluHis 80
Db 645 CTTTCTTAATAAATACATCCAGAGGTCTCCACCATGAGTGGGAAAGCCATAGAGCAT 704
Qy 81 LeuAlaIlySGInGlyAsnArgPheHisPheAspIlelyProProLeuHisHisAlaIly 100
Db 705 TTGGCCAAACAGAGAAATAGATTCATTTTGCATCAAACTCCCTGCATCATCTAA 764
Qy 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleMetIlyIlyS 120
Db 765 AATTGTATTTTTCTTTACTGAGCTTCAACCGTTACTGATTAATAATATGAAAAAG 824
Qy 121 GluIlySGInGlyIle-----GluIlySGInIleleuSerAla 126
Db 825 GAAAAAGGAAAGGATATTTCTAATAGTAAGTTGAAAGACATAAATATTCCTGATG 884
Qy 127 -----GluIlySGInIleleuSerAla 135
Db 885 TGCCTAAAAATAGCTCTCATTTCTGACGATAGAAAGGGGCAATCTCTCTCAGCA 944
Qy 136 AlaAspIleAlaIleThrValGlnHisThrMetAlaCyHisIleValIlyArgThrHis 155
Db 945 GCAGACATTCCTGCCACAGTACAGCACAAATGGCATGTCATCTGTGAAAAAGACACT 1004
Qy 156 ArgAlaIleleuPheCyIlySGInArgAspLeuLeuProGlnAsnAsnAlaValIleVal 175
Db 1005 CCGGCTATTCCTGTTTGTGAACAGAGACATTTGTTACTCAAAATAATGCACTAGTGT 1064
Qy 176 AlaSerGlyIlyValAlaSerAsnPheTyrlleArgArgAlaLeuGlnIleleuThrAsn 195
Db 1065 GCATCTGTGTGTGTGCGCAAGTAACTTCTATATCCGCAAGCTCTGAAATTTTAAACAAC 1124
Qy 196 AlaThrGlnCySerThrLeuLeuCySerProProAlaGluCyThrAspAsnGlyIleMet 215
Db 1125 GCACACAGAGTCACTTGTGTGTCTCTCCACACATGATGCAAGTAATGCGCTTTATG 1184
Qy 216 IleAlaIleArgPheGlyIleGluArgLeuArgIleGlyIleLeuHisAspIleGlu 235
Db 1185 ATTCATGAGATGATGATTTGAAAGACTACGTGCTGTGGCATTTTACATGACATAGAA 1244
Qy 236 GlyIleArgTyrgIupProIlyCySerProleuGlyValAspIleSerIlyGlyValGlyGlu 255
Db 1245 GCGATCGGCTATGAAACCAAAATGCTCTTGAGATGACATATCAAAAGAGTTGAGAA 1304
Qy 256 AlaSerIleIlyValProGlnLeuIlyMetGluIle 267
Db 1305 GCTTCCATAAAAGTACCAATTAATAATGAGATTA 1340
RESULT 6
HSA295148 1387 bp mRNA linear PRI 30-OCT-2000
LOCUS Homo sapiens mRNA for putative sialoglycoprotease type 2.
DEFINITION AJ295148.1 GI:11071726
VERSION sialoglycoprotease type 2.
KEYWORDS metalloproteinase; sialoglycoprotease.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Chen,J.M., Fortunato,M. and Barrett,A.J.
TITLE Cloning and sequencing of a second human putative
JOURNAL sialoglycoprotease homologue
AUTHORS Unpublished
REFERENCE 2 (bases 1 to 1387)
AUTHORS Chen,J.M.
TITLE Direct Submission
JOURNAL Submitted (27 OCT-2000) Chen J.M., McC Molecular Enzymology
AUTHORS Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT,
FEATURES UNITED KINGDOM
Location/Qualifiers
```

```
source 1. 1387
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/feature_type="utermus"
24. 1343
/function="metalloproteinase of family M22"
/codon_start=1
/product="putative sialoglycoprotease type 2"
/protein_id="CAC14666.1"
/db_xref="GI:11071727"
/db_xref="GOA:Q9H4B0"
/db_xref="UniProt/TREMBL:Q9H4B0"
/translation="MLITKTAGVFPKPSKRYEFLASFNPEPTLFLAKIVLGIST
SCDDPAVAVDEKGVNLTGEAIIHQTEVHLTKGTGIVPAQAQIHRNIQRIQVEALAS
GVSPSDLSAIIATIKRGALSLGVLSFSLVGLGKKEPIPIHHEAVALFRLTNK
VAFPIVLIIISGHCILALVGVSDFLGKSLIDAPGMILKVARSLIKPEGST
MSGKRAIEHLAQGRFHDIPRLHAKRCPSFGLQVHTDKIMKEKEGIFLI
SRVEQINIFGLCLKIAHFCRYEKGLISSADIIATVQHTACHLVKRTHTAIFLCK
ORDLIPQNNVAIVASGVASNFYIRRALIILTNATQCTLCPEPRCLTDNGIMIAWNG
IERLKGILIHIDIGIRYERKPLGVDSKEVGEASIRVPLQMEI"
ORIGIN
Alignment Scores:
Pred. No.: 5,43e-124 Length: 1387
Score: 1358.50 Matches: 266
Percent Similarity: 91.10% Conservative: 0
Best Local Similarity: 91.10% Mismatches: 1
Query Match: 98.09% Indels: 25
DB: 9 Gaps: 1
US-10-649-273-2_COPY_148_414 (1-267) x HSA295148 (1-1387)
Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIlyValGluPheProPheLeuVal 20
Db 465 ATGAGAGGCTCAGTACCTTAATTAAGTTGACCAAAATGATTAATTTCTTTTATGTT 524
Qy 21 LeuLeuIleSerGlyGlyHisCySerLeuLeuAlaLeuValGlnIlyValSerAspPheLeu 40
Db 525 CTTTGTATTTCTGAGAGTACCTGTCTGTGGCATTTGTTCAAGAGATTTCAAGATTTTCTG 584
Qy 41 LeuLeuGlyIlySerLeuAspIleAlaProGlyAspMetLeuAspIlyValAlaArgArg 60
Db 585 CTTCTGGAAAGTCTTTGGACATAGCACAGGATGATCTTGACAAAGTGGCAAGAA 644
Qy 61 LeuSerLeuIlelyHisProGluCySerThrMetSerGlyLyalaIleGluHis 80
Db 645 CTTTCTTAATAAATACATCCAGAGGTCTCCACCATGAGTGGGAAAGCCATAGAGCAT 704
Qy 81 LeuAlaIlySGInGlyAsnArgPheHisPheAspIlelyProProLeuHisHisAlaIly 100
Db 705 TTGGCCAAACAGAGAAATAGATTCATTTTGCATCAAACTCCCTGCATCATCTAA 764
Qy 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleMetIlyIlyS 120
Db 765 AATTGTATTTTTCTTTACTGAGCTTCAACCGTTACTGATTAATAATATGAAAAAG 824
Qy 121 GluIlySGInGlyIle-----GluIlySGInIleleuSerAla 126
Db 825 GAAAAAGGAAAGGATATTTCTAATAGTAAGTTGAAAGACATAAATATTCCTGATG 884
Qy 127 -----GluIlySGInIleleuSerAla 135
Db 885 TGCCTAAAAATAGCTCTCATTTCTGACGATAGAAAGGGGCAATCTCTCTCAGCA 944
Qy 136 AlaAspIleAlaIleThrValGlnHisThrMetAlaCyHisIleValIlyArgThrHis 155
Db 945 GCAGACATTCCTGCCACAGTACAGCACAAATGGCATGTCATCTGTGAAAAAGACACT 1004
Qy 156 ArgAlaIleleuPheCyIlySGInArgAspLeuLeuProGlnAsnAsnAlaValIleVal 175
Db 1005 CCGGCTATTCCTGTTTGTGAACAGAGACATTTGTTACTCAAAATAATGCACTAGTGT 1064
```

QY 176 AlaSerGlyValAlaSerAspNheThrIleArgAlaLeuGluIleLeuThrAsn 195  
DB 1065 GCATCTGGTGGTGGCAAGTAATTCATATCCGACAGCTCTGGAATTTTAAACAAC 1124  
QY 196 AlaThrGlnCysThrIleLeuCysProProAlaGluCysThrAspAsnGlyIleMet 215  
DB 1125 GCAACACAGGACCTTGTGTGTCTCTCCGACACATGACATGAAATGCGATTATG 1184  
QY 216 IleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235  
DB 1185 ATTGCATGGAATGGTATTTGAAAGACTACGTGGCTTGGGCAATTTTACATGACATGAA 1244  
QY 236 GlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGluValGlyGlu 255  
DB 1245 GGCATCCGCTATGAAACCAATGTCCTTGGAGTAGACATATCAAAAGAGTTGGAGAA 1304  
QY 256 AlaSerIleLysValProGlnLeuLysMetGluIle 267  
DB 1305 GCTTCATAAAGTAACCAATTAATAATGAGATTA 1340

RESULT 7  
AX664697 1245 bp DNA linear PAT 22-MAR-2003  
LOCUS AX664697 Sequence 6 from Patent WO02074960.  
ACCESSION AX664697  
VERSION AX664697.1 GI:29164457  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1  
AUTHORS Leiby K.R., Kapeller-Libermann, R. and Glucksmann, M.  
TITLE 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions  
JOURNAL of human proteins and uses thereof  
Patent: WO 02074960-A 6 26-SEP-2002;  
Millennium Pharmaceuticals, Inc. (US)  
FEATURES  
source 1. 1245  
Location/Qualifiers  
1. 1245  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Alignment Scores:  
Pred. No.: 5.14e-123 Length: 1245  
Score: 1348.00 Matches: 260  
Percent Similarity: 98.50% Conservative: 3  
Best Local Similarity: 97.38% Mismatches: 4  
Query Match: 97.33% Indels: 0  
Gaps: 0  
DB: 6

US-10-649-273-2\_COPY\_148\_414 (1-267) x AX664697 (1-1245)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
DB 442 ATGAGGCTCATGCACTTACTATTGAGTTGACCAATAAGTGAATTTCTTTTATGTT 501  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 502 CTTTGTGATTTCTGGAGGTCACTGCTGTGGCATTTAGTTCAAGGATTTTCAATTTTCTG 561  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
DB 562 CTTTGTGAAAGTCTTTTGACATACACACGCTGACATGCTTGACACAGGTGGCAAGAA 621  
QY 61 LeuSerLeuIleLysHisProGlnCysSerThrMetSerGlyGlyLysAlaIleGluHis 80  
DB 622 CTTTCTTTAAATAAAACATCCAGAGTCTCACCAATGAGGTGGGAAAGCCATGAAACT 681  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
DB 682 TTGGCCAAACAGAGAAATGATTTCAATTTTGACATCAACCTCCCTTGACATATGCTAAA 741

QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
DB 742 AATTGGATTTTCTTTTACTGCACTTCAACAGCTTACATGATTAATAATGAAACACAG 801  
QY 121 GlyLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
DB 802 AAACAAGAGGAAGGATTTGAGAAAGGGCAAACTCTGTCTTCAGCAGACACATTTGCTGC 861  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
DB 862 ACGATACACACACAAATGCAATGTCTCTTGGAAAAAGAACACATGGGCTATTTCTGTT 921  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyVal 180  
DB 922 TGTAAAGAGAGACTTGTATCTTCAATAATATGCAGTACTGGTTCATCTGGTGGTGC 981  
QY 181 AlaSerAspNheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
DB 982 GCAGATTAATCTTATATCCGACAGCTCTGAAATTTTAAACAACCAACACAGTGCACCT 1041  
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
DB 1042 TTGTTGTGTCCTCTCCGACACTATGACATGATTAATGCAATTAATGATGCAATGGT 1101  
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240  
DB 1102 ATTGAAAGCTACGTCCTGGCTTGGGCAATTTTACATGACATGAAAGCATCCGCTATGAA 1161  
QY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260  
DB 1162 CCAAAATGTCCTTGGAGTAGACATATCAAAAGAGTTGGAGAACTTCATTAATAAGTA 1221  
QY 261 ProGlnLeuLysMetGluIle 267  
DB 1222 CCACATTAATAAATGAGATTA 1242

RESULT 8  
AX664695 1820 bp DNA linear PAT 22-MAR-2003  
LOCUS AX664695 Sequence 4 from Patent WO02074960.  
ACCESSION AX664695  
VERSION AX664695.1 GI:29164455  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1  
AUTHORS Leiby K.R., Kapeller-Libermann, R. and Glucksmann, M.  
TITLE 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions  
JOURNAL of human proteins and uses thereof  
Patent: WO 02074960-A 4 26-SEP-2002;  
Millennium Pharmaceuticals, Inc. (US)  
FEATURES  
source 1. 1820  
Location/Qualifiers  
1. 1820  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
146. 1390  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD80044.1"  
/db\_xref="GI:29164456"  
/translation="MLILTKTAGVFPFKRKRYEFLRSFNFHPGTLFLKIVLGLET  
SCDDPRAAVDEGTGNVIGRAIHSGTEVNHKTGIVIPPAQOLHRENIORVQALASAS  
GSPSDLSAIVATTIKGALSIGVGSFSLQVIGQAKKPIPIHHEAHMLITRLTNK  
VSPRPVLIISGHCLIALVQGVSDFLIKGSLIDIPGMLEDAVARLSLTKHPEST  
MSGKRAIEHLAQRNFHDIDKPLHAKKCDPSFGLQHTYTDKNENRQDEGIEKG  
QILSSADIAATVQHTWACHLVKTRTRAILPFCKQRLDPQNNAVLVAAGVANSFTIR  
RALBITVNAVQCTLCPPRLCTDNGIMTAMWNGIEIRLRAIGLIDHIEGIRYKPCPL  
GVDISKEVEASIKIVQLKMEI"

ORIGIN









George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/IMLUL at: <http://image.llnl.gov>  
Series: IRAK Plate: 86 Row: f Column: 12.

## FEATURES

## Source

Location/Qualifiers  
1..1017  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5053559"  
/issue\_type="Liver, normal, 5 month old male mouse."  
/clone\_id="NCI CGAP\_L19"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

## ORIGIN

## Alignment Scores:

Pred. No.:	6,53e-111	Length:	1017
Score:	1224.00	Matches:	233
Percent Similarity:	92.51%	Conservative:	14
Best Local Similarity:	87.27%	Mismatches:	20
Query Match:	88.38%	Indels:	0
DB:	10	Gaps:	0

US-10-649-273-2\_COPY\_148\_414 (1-267) x BC038910 (1-1017)

```
OY      1 MetGluAlaHlaAleuthrIlleargLeuthrAsnlyValGluPheProPheleuVal 20
DB      26 ATGAGAGCTCAGCAGCTGACTATTAGGCTCAACATAAGATGAAATTCCTTTTAACT 85
OY      21 LeuLeuIleSerGlyVhiisCyaleuLeuAlaLeuValGlnGlyValSerAspPheleu 40
DB      86 CTTTGGATTCTGGCGGCTGCTGCTGTGGCATTAATCAAGGTTTCGAGTTCCG 145
OY      41 LeuLeuGlyLySerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 60
DB      146 CTCCTTGGAAAGTCTTGGACATAGCACAGGCGACAGCTTGGACAAGTGGCAAGAGA 205
OY      61 LeuSerLeuIleLyshisProGlyCySerThrMetSerGlyGlyValAlaIleGluHis 80
DB      206 CTTTCTTAATCAAAACATCCAGAAATGTTCTCAATGAGTGGGAAAGCTATAGAACAG 265
OY      81 LeuAlaLySGlnGlyAsnArgPheHisPheAspIleLyProProLeuHisHlaAlaLyS 100
DB      266 TTGGCCAAAGACGGAATAGATTCATTTCATCAATCAACCTATGACAGATCTAG 325
OY      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLySIlleMetLySlys 120
DB      326 AATTGCAATTTTCTTTCACGGGACTTCAACATATTCTGATAGCTAATTAACACACAG 385
OY      121 GluLySGlnGlyIleGluLySGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB      386 GAAAGAAAGAGGACATAGAGAGGGCAAAATCCGTATCAGTCGAGACATTCCTCT 445
OY      141 ThrValGlnHisPheMetAlaCyehisLeuValLySArgThrHisArgAlaIleLeuPhe 160
DB      446 GCGGTACAGCATGCAACAGCGCTGCACCTTCGAAAGAACACATCGCGCTATTCGTTT 505
OY      161 CyaleuGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180
DB      506 TGCAGAGGAAATAATTTGCTCTCTTCAGCTAACGAGATTAAGTTGATTCGAGAGCTT 565
OY      181 AlAserAsnPheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCyThr 200
DB      566 GCAAGTAATCTTGATCCGAAAGACATTCGAAATTCGCAAAATGCAACGACAGTGCACG 625
OY      201 LeuLeuCySProProArgLeuCyThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
DB      626 TTGTGTGTCCACTCCAGACCTGTCGACTGACATGCAATGCAATGATTCGATGGAATGGA 685
OY      221 IlleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrlu 240
```

```
DB      686 ATTGAAGATTAAGTCTGCGCTTGGCTTTTAAACATGCTAGAGACATCCATATGAA 745
```

```
OY      241 ProLyCySProLeuGlyValAspIleSerLyGlyValGlyGluAlaSerIleLySVal 260
DB      746 CCAAAATGCTCTTTCGAGTAGACATATCCAGAGAAAGTTGCAGAACTGCCATTAAGTA 805
```

```
OY      261 ProGlnLeuLyMetGluIle 267
DB      806 CCGCATTAATAATGCACTT 826
```

## RESULT 11

AX713716  
LOCUS AX713716  
DEFINITION Sequence 400 from Patent EPI293569.  
ACCESSION AX713716  
VERSION AX713716.1 GI:29888642

## KEYWORDS

## SOURCE

## ORGANISM

Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J., I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuko, Y.

Full-length cDNAs  
Patent: EP 1293569-A 400 19-MAR-2003;  
Helix Research Institute (JP) ; Research Association for  
Biotechnology (JP)

## FEATURES

source 1..2208  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## ORIGIN

## Alignment Scores:

Pred. No.:	1.57e-108	Length:	2208
Score:	1204.00 <td>Matches:</td> <td>239</td>	Matches:	239
Percent Similarity:	89.51% <td>Conservative:</td> <td>0</td>	Conservative:	0
Best Local Similarity:	89.51% <td>Mismatches:</td> <td>4</td>	Mismatches:	4
Query Match:	86.93% <td>Indels:</td> <td>24</td>	Indels:	24
DB:	6 <td>Gaps:</td> <td>1</td>	Gaps:	1

US-10-649-273-2\_COPY\_148\_414 (1-267) x AX713716 (1-2208)

```
OY      1 MetGluAlaHlaAleuthrIlleargLeuthrAsnlyValGluPheProPheleuVal 20
DB      785 ATGAGAGCTCAGCAGCTTTCATTAAGTTCGACCAATTAAGTACAAATTTCTTTTAACT 844
OY      21 LeuLeuIleSerGlyVhiisCyaleuLeuAlaLeuValGlnGlyValSerAspPheleu 40
DB      845 CTTTGGATTCTGAGAGTCACTGTCGTTGGGATTAAGTTCAAGAGATTCAGATTTTCG 904
OY      41 LeuLeuGlyLySerLeuAspIleAlaProGlyAspMetLeuAspLySValAlaArgArg 60
DB      905 CTTTCTTAATCAAAACATCCAGAAATGTTCTCAATGAGTGGGAAAGCTATAGAACAG 964
OY      61 LeuSerLeuIleLyshisProGlyCySerThrMetSerGlyGlyValAlaIleGluHis 80
DB      965 CTTCTTAATAATCAATCAAGAGTCTCACCATAGAGTGTGGAAAGCCATAGAACAT 1024
OY      81 LeuAlaLySGlnGlyAsnArgPheHisPheAspIleLySProProLeuHisHlaAlaLyS 100
DB      1025 TTGGCCAAAGACGAAATAGATTCATTTGACATCAAACTCCCTTGACATATGCTAAA 1084
OY      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLySIlleMetLySlys 120
DB      1085 AATTGGAATTTCTTTTACTGCACTTCAACAGTTAAGATTAATAATTAAGAAAAG 1144
OY      121 GluLySGlnGlyIleGluLySGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
```

Db	Accession	Version	Source	Organism	Reference Authors
Db	1145	AK05441	LOCUS DEFINITION	AK05441	1
Qy	141	AK05441	LOCUS DEFINITION	AK05441	1
Db	1205	AK05441	LOCUS DEFINITION	AK05441	1
Qy	161	AK05441	LOCUS DEFINITION	AK05441	1
Db	1265	AK05441	LOCUS DEFINITION	AK05441	1
Qy	181	AK05441	LOCUS DEFINITION	AK05441	1
Db	1325	AK05441	LOCUS DEFINITION	AK05441	1
Qy	201	AK05441	LOCUS DEFINITION	AK05441	1
Db	1385	AK05441	LOCUS DEFINITION	AK05441	1
Qy	221	AK05441	LOCUS DEFINITION	AK05441	1
Db	1435	AK05441	LOCUS DEFINITION	AK05441	1
Qy	241	AK05441	LOCUS DEFINITION	AK05441	1
Db	1436	AK05441	LOCUS DEFINITION	AK05441	1
Qy	261	AK05441	LOCUS DEFINITION	AK05441	1
Db	1493	AK05441	LOCUS DEFINITION	AK05441	1
RESULT 12					
AK05441					
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					

TITLE	Complete sequencing and characterization of 21,243 full-length human cDNAs					
JOURNAL	Nat. Genet.	36	(1),	40-45	(2004)	
PUBMED	14702039					
REFERENCE	2					
AUTHORS	Nishii,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H., Sugiyama,T., Irie,R., Otuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Yamachi,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanba,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuno,Y., Nagai,K. and Isogai,T.					
TITLE	NEBO human cDNA sequencing project					
JOURNAL	Unpublished					
REFERENCE	3 (bases 1 to 2208)					
AUTHORS	Isogai,T., Otuki,T. and Sugiyama,T.					
TITLE	Direct Submission					
JOURNAL	Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomic@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)					
COMMENT	NEBO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology (RAB); cDNA library construction; Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing; RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.					
FEATURES	Location/Qualifiers					
SOURCE	1..2208					
	/organism="Homo sapiens"					
	/mol_type="mRNA"					
	/db_xref="taxon:9606"					
	/clone="FEBRA2004592"					
	/tissue type="brain"					
	/clone_id="FEBRA2"					
	/dev_stage="fetus"					
	/note="Cloning vector: pME18SFL3"					
CDS	344..1438					
	/note="unnamed protein product"					
	/codon_start=1					
	/protein_id="BAB70923.1"					
	/db_xref="GI:16550167"					
	/translation="MLLIRKTAGVFPSPSKRYTERLRSNPHDGLFLHKIVGISTSCDDTAAVVDGTGVNGAIIHSQTEVALTKGGIVPPAAOQLHEENIQRLVEELSSASGVSDLSAIATTKIKPLALSLVGSLFSIQLVOGLKKPPIPIHMEAHALLRLTKRVPEPVLISIGSGCLIALVOVGDFLLDKSLDIAGDMLDVARLPILKHCESTMSGCAKEHLAKCGNRHPDIKPPIHHAKNCDDPFGTLOHYTDKIINKKKEBIEKGOLISSADIAATYQHTACHLVYKRTRAILPFCKORDLLPNNNAVVASGVASNFCIRPALEILNATQCTLCPPLRCTDNGIMIA"					
ORIGIN						
Alignment Scores:						
Pred. No.:	1.57e-108	Length:	2208			
Score:	1204.00	Matches:	239			
Percent Similarity:	89.51%	Conservative:	0			
Best local Similarity:	89.51%	Mismatch:	4			
Query Match:	86.93%	Indels:	24			
DB:	9	Gaps:	1			
US-10-649-273-2_COPY_148_414	(1-267) x AK055441 (1-2208)					
Oy	1 MetGuaAaIaIaaleuThrlleaargLeuThraAnlysaValgiupheProPhelauVal20					
Dd	785 ATGAGAGGTACATGCATTACTATTAGTTGACCAATAAGAATTCCTTTTATGTT844					
Oy	21 LeuleuIlseSerglglYhIsCyaleuLeuAlalaueValginglYvalSerAspheleu40					
Dd	845 CTTTGATTTCGAGAGTCACCTGCTGTGGCATATTGTCAAAGAGTTTCAGATTTCG904					
Oy	41 LeuleuGlYysSerrleuAspIlalalPRoglYasphMeleuAspLyvsvalAlaArg60					
Dd	905 CTCTTGGAAGCTCTTGACAATGACCCAGAGTACATGCTTGACAGAGTGGCAGAAGA964					

Oy		61	LeuSERleuLILeYSHiSPiPrGILCYSeRThrMeTSeRGYLgLYVAlIIGeLUIS	80
Db		965	CTTCCTTTAAATAAACATCCAGAGCGCTCCACCATGACTGGTGGAAAGCCATTGAACAT	1024
Oy		81	LeuAIALYSGINGlYASnaRYPhHeHISphEAsPILeLYSProProLeuHiHIIAlALYS	100
Db		1025	TTGGCCCAACAAGGAAATAAGATTTCATTTCATTTCATCAACCTCCCTTGCAATCATGTCAA	1084
Oy		101	ASnCYsaSPheSeRePhEThrGlyVLengInHISValTHraBPLYSIlElIMeTLeyLS	120
Db		1085	AATTGTGATTTTCTCTTACTGACCTTAACACGTTACTGATTAATAATATATGAAAAAG	1144
Oy		121	GluYSGIngUlgUgYllEGlULySGlYGlNILEuSeRSerLaIAIASPIleAla	140
Db		1145	GAABAAAGAGAAGGATATTGAGAGGGCGAAATTCGTCTTCAGCAGCAGACATTGCTGCC	1204
Oy		141	ThrVALglnHISThrMeTAlCYSHISleuVALylYBaGrTHRHaRgAlAILeUPHe	160
Db		1205	ACACTTACGCAGCACAAATGCGATGTCATCTTGTGAAAAMAAACATCCGGCTATTCTGTTT	1264
Oy		161	CysLYSGINAgaAPLeuLeuProGLnaSmAmnAlAVAlleuVALIASerGlyGLYAl	180
Db		1265	TGTAAAGCAGAGAGACTTGTACTTCCTCAATATATGACGATCTGTTGCAATCTGGTGTCTC	1324
Oy		181	AlaseRaSnPhetyRIleArgARglalENGulileuThrrASnaAlATHGInCYeThr	200
Db		1325	GCAAGTAACTTCGTATCCGCGAGAGCTCTGAAATTTTAACAAAGCAGACAGTGCACCT	1384
Oy		201	LeuLeuCYPProProProARGLeUCYthrASpangilylleMETIlaALTTPsangly	220
Db		1385	TTGTTGTGTCTCCTCCGCCAGCATATGACCTATATGCAATTATGATTCATTCGA-----	1435
Oy		221	IlEGlUlArGLEuARGAlagLYleugLYllEleuHIASpILIEglUGlylIleArgTYcLU	240
Db		1435	-----	1435
Oy		241	ProLYSCYPProLeUGLYValASPILesERLySGlUValGlyUALaserIlElySVal	260
Db		1436	---TGATGTCTCTTTGGAGTGAACATCAAAAAGAGTTGAGAAAGCTTCATTAAAGTA	1492
Oy		261	ProGLnuLeuLYMetGLuIlle	267
Db		1493	CCACAATTAAAAATGAGATA	1513
RESULT_13				
BC078974				
LOCUS		BC078974	1546 bp	mRNA linear ROD 03-AUG-2004
DEFINITION		Rattus norvegicus cDNA clone IMAGE:7111906,	partial cds.	
ACCESSION		BC078974		
VERSION		BC078974.1	GI:50926879	
KEYWORDS		.		
SOURCE		Rattus norvegicus (Norway rat)		
ORGANISM		Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
		Rattus.		
REFERENCE		1 (bases 1 to 1546)		
AUTHORS		Strausner,R.L., Feingold,R.A., Grouse,L.H., Derge,J.G.,		
		Klausner,R.D., Collins,F.S., Wagner,L., Shemen,C.M., Schuller,G.D.,		
		Altshuler,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,		
		Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heiseh,P.,		
		Ditchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,		
		Steidlton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,		
		Schreier,T.E., Brownstein,M.J., Udell,T.B., Toobyanki,S.,		
		Carninci,P., Prange,C., Kaha,S.S., Loquellano,N.A., Peters,G.J.,		
		Abrahamson,R.D., Mulihay,S.J., Bosak,S.A., McEwan,P.J.,		
		McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,		
		Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyar,S.W.,		
		Villaloun,D.K., Muzny,D.M., Sodergren,K.J., Lu,X., Gibbs,R.A.,		
		Faney,J., Helton,E., Ketteman,M., Madan,A., Rodriguez,S.,		
		Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,		
		Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,B.D.,		

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1  
2 (bases 1 to 1546)  
Director MGC Project.  
Direct Submission  
Submitted (02-AUG-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Howard Jacobs  
CDNA Library Preparation: Express Genomics  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www.sbgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

REMARK  
COMMENT

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ULNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 182 Row: f Column: 6  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.  
Location/Qualifiers  
1. 1546  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
/clone="IMAGE:7111906"  
/tissue\_type="Testis, rat (Brown Norway)"  
/clone\_id="NIH\_MGC\_237"  
/lab\_host="DH10B"  
/note="Vector: pExpress1"  
144. >1546  
/codon\_start=1  
/product="unknown (protein for IMAGE:7111906)"  
/protein\_id="AAH8974.1"  
/db\_xref="GI:50926880"  
/translation="MLMTSKTAGIIPPPRSNVRGPIRRFNQVRPALFHHKLVIGTSCDDTAAAVVDERTGNVIGKALHSTGTEVPHATKCGIVPPVAQJLRNIOQRYEALSNASGVSSDSIAIATTIKPGLALSLGGLSFSVOLNVEKPTPIHMEAHALVTLRLTKVGEPFVILISGHCILALVOSVDFLLAKCSDIDIAQMDLKVARRLSIKHPECSYMSGGKAIEHLAKENRHFPTINPMOVAKKCDSPFGIQHTVDKLITKKEKEGIEKGQIISSADIAAAVOHATACGACAKRTBAIIFCCOQKINSANNAVLVSGVAKLVYIRRALEIVANATOCITLCPEPLCTDNGIMIMANGIERLRALGILHVEDIVRPNKGEIVSKLRVLTALTEDLDVSSVTHVAHSPANSGRANITQTSKSCYQTYMRTVRRITLNIIEKSKKKKKKKKKK"

ORIGIN

Alignment Scores:  
Pred. No.:  
Score:  
Percent Similarity:  
Best Local Similarity:  
Query Match:  
DB:

Pred. No.	Score	Percent Similarity	Best Local Similarity	Query Match	DB
4.7e-100	1116.00	92.56%	88.43%	10	US-10-649-273-2_COPY_148_414 (1-267) x BC078974 (1-1546)
Length: 1546	Matches: 214	Conservative: 10	Mismatches: 18	Indels: 0	Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x BC078974 (1-1546)

OR  
DB

1  
585

MeG1u1aht1sa1a1eThr1leArg1eThra1n1yVa1G1uPh1eProPh1eUa1 20  
|||||:::|||||  
ATGAGGCTACGCGCTGACATATTAGGCTACCAACAAGTCGATTCTTTTATAGTT 644

QY 21 LeuLeu11eSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 645 CTTTGTGATTTCTGAGAGCCACCTGCTGTGGCGTTAGTTCAAGAGTTTCAGATTTCG 704  
QY 41 LeuLeuGlyLysSerLeuAsp11eAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
DB 705 CTCTCGGGAAGTCCCTGAGACATAGCGCAGCGCACATGCTTGACAAAGGTGCAAGAA 764  
QY 61 LeuSerLeu11eLysHisProGlyLysSerThMetSerGlyGlyValAla11eGluHis 80  
DB 765 CTTTCTTTAATCAAAATCCAGAAATGTTCTACAAATGATGGTGGGAAAGCTTAAACAT 824  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisSAspAsp11eLysProProLeuHisHisAlaLys 100  
DB 825 TTGGCCAAAGAAAGAAATAGATTCCACTTACTACTATCAATCCACCACAGAAATGCTAG 884  
QY 101 AsnGlyAspPheSerPheThGlyLeuGlnHisValThrAspLys11eLeuMetLysLys 120  
DB 885 AACTGTATTTTCTTTTACGCGACTTCAACATGTCACCGATMACCTAATAACACACAAAG 944  
QY 121 GluLysGlnGlyGlyLeuGlyGln11eLeuSerSerAlaAlaAsp11eAlaAla 140  
DB 945 GAAAGAAAGAAAGCATTTGAGAAAGGCGAAATCTGTATCAAGCCGACATTCCTGCT 1004  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAla11eLeuPhe 160  
DB 1005 GCGGTACAGACAGCAACAGCGGTGCCACTTGCGAAAGAAACATCGTCTATTCGTTT 1064  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
DB 1065 TGCACAGCAAGAAATTTGCTATCTTCACGATTAACGATTAATGTTGTCTGAGAGTGT 1124  
QY 181 AlaSerAsnPheTyrl1eArgArgAlaLeuGln11eLeuThrAsnAlaThrGlnCysThr 200  
DB 1125 GCAAGTAATCTTGATCATCGAAGACATTTGAAATTTAGCAAAATGCAACAAATGCACT 1184  
QY 201 LeuLeuCysProProProArgLeuLeuCysThrAspAsnGly11eMet11eAla11eTrpAsnGly 220  
DB 1185 TTGTTGTGTCCCTCCGAGACTGTGCACTGACATGATGATATGATTCAGATGCAATGCA 1244  
QY 221 T11eGluArgLeuArgAlaGlyLeuGly11eLeuHisAsp11eGlnGly11eArgTyrglu 240  
DB 1245 ATTGAAGATTAACGTGCTGGCTGGCAATTTTACATGATGATGAAAGACATCCGATACGA 1304  
QY 241 ProLys 242  
DB 1305 CCAAG 1310  
RESULT 14  
BC051211 1109 bp mRNA linear ROD 15-APR-2003  
LOCUS Mus musculus, clone IMAGE:1327545, mRNA.  
DEFINITION BC051211  
ACCESSION BC051211.1 GI:29881634  
VERSION  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1109)  
Strausberg, R.  
Direct Submission  
Submitted (14-APR-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIR-MGC Project URL: http://mgc.nci.nih.gov  
Contact: MGC help desk  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Marcello Bento Soares, Ph.D.  
Library Preparation: M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
DNA Sequencing by: Institute for Systems Biology

http://www.systembiology.org  
contact: amadan@systemsbiology.org  
Anup Madan, Jessica Fahy, Erin Halton, Mark Ketteman, Anuradha  
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov  
Series: IRAX Plate: 113 Row: b Column: 1  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis.  
FEATURES  
location/Qualifiers  
1..1109  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1327545"  
/tissue\_type="Thymus gland, mouse"  
/clone\_11b="Soares\_thymus\_2NBMt"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac"  
ORIGIN  
Alignment Scores:  
Pred. No.: 6.23e-84 Length: 1109  
Score: 950.50 Matches: 190  
Percent Similarity: 79.10% Conservative: 22  
Best Local Similarity: 70.90% Mismatches: 35  
Query Match: 68.63% Indels: 21  
DB: 10 Gaps: 4  
US-10-649-273-2\_COPY\_148\_414 (1-267) x BC051211 (1-1109)  
QY 18 PheLeuValLeuLeu11eSer-----GlyGlyHisCysLeu----- 29  
DB 82 TTTTGTGATTTCTGATTCGATTCGCTGATTCGCTGCAATTTTGAACATCCATATCTCT 141  
QY 30 -----LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAsp 47  
DB 142 TCTGTTTAAACTGACTTGCTGCATATGAGGTTATCTTTGTTATAT-----AAT 192  
QY 48 T11eAlaProGlyAspMetLeuAspLys-----ValAlaArg 59  
DB 193 TTGCAGCCAAATATATTTATGAAAGAAATGTAATGTTTAAATAGTGGCAAG 252  
QY 60 ArgLeuSerLeu11eLysHisProGlyLysSerThMetSerGlyGlyValAla11eGlu 79  
DB 253 AGACTTTCTTTAATCAAAATCCAGAAATGTTCTACATGATGTTGTGGAAGAAAGCTATAGAA 312  
QY 80 HisLeuAlaLysGlnGlyAsnArgPheHisSAspAsp11eLysProProLeuHisHisAla 99  
DB 313 CAGTTGGCCAAAGACGAAATGATTCATTTTACTATCAATCACCTATGCGCAATGCT 372  
QY 100 LysAsnGlyAspPheSerPheThGlyLeuGlnHisValThrAspLys11eLeuMetLys 119  
DB 373 AAGAAATGCGATTTTCTTCAAGGCACTTCAACATATTAAGTATTAAGTATTAAGCAAC 432  
QY 120 LysGluLysGlnGlyGlyLeuGlyGln11eLeuSerSerAlaAlaAsp11eAla 139  
DB 433 AAGGAAAGAAAGAAAGGCAATGGAAGGCGCAAAATCTGTCAATCACTGACAGCAATGCT 492  
QY 140 AlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAla11eLeu 159  
DB 493 GCTGCGGTACAGACAGCAACAGGTGCCACTTGCGAAAGAAACACATGCGGTATTCG 552  
QY 160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGly 179  
DB 553 TTTTGCACAGCAAGAAATTTGCTCTCTCCAGCTTAAGCGCATATTAAGTATTCGAGCT 612  
QY 180 ValAlaSerAsnPheTyrl1eArgArgAlaLeuGln11eLeuThrAsnAlaThrGlnCys 199  
DB 613 GTTGCAGATTAATCTGATCAATCCGAAAGCAATTCGAAATTCGCAATGCAAGCAAGTGC 672

QY	200	ThrinleuleusCySPROProProArgueneCyPhrArpaPaengIyleMeTlialaTrPhen	213
Db	673	ACGTTGTGTGTGCACCTCCCAAGATGTGGCATGCAAAAGGATCATATATTCACATGAAT	732
QY	220	GlylliegIuaNgLeuAraGalagIyLengIylleUeuhIaaPilegugIyIleaRgtYr	239
Db	733	GGAATTGAAAGATTACGTGCTGGCGCTTTTACATGATGATGAAGAATCCGATAT	792
QY	240	GIUpPolyesCyEProLeuGIyValAspIleSerIysglIyValGIyGlunAlaserIleyS	259
Db	793	GAACCAAATATGCTCTTGAGTAGACAATATCCAGAAAGTTGCAGAAAGTCGCAATAAA	852
QY	260	ValProGInleUlysMetCguIle	267
Db	853	GTACCGCGATTAAAAATGCACTT	876
RESULT 15			
LOCUS	BX934991	1522 bp	mRNA linear VRT 02-FEB-2004
DEFINITION	Gallus gallus finished cDNA, clone CHEST189114.		
ACCESSION	BX934991		
VERSION	BX934991.1	GI:41635519	
KEYWORDS			
SOURCE			
ORGANISM	Gallus gallus (chicken)		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Archaeosauria; Aves; Neognathae; Galliformes; Phasianidae;		
	Phasianinae; Gallus		
	1 (bases 1 to 1522)		
REFERENCE	Boardman,P.E., Bonfield,J.K., Brown,W.R.A., Carder,C., Chalk,S.E.,		
AUTHORS	Croning,M.D.R., Davies,R.M., Francis,M.D., Gatham,D.V.,		
	Hubbard,S.J., Humphray,S.J., Hunt,P.J., Maddison,M., McLaren,S.R.,		
	Niblett,S., Overton,I.M., Rogers,J., Scott,C.E., Taylor,R.G.,		
	Tickle,C. and Wilson,S.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,		
	CB10 1SA, UK. E-mail enquiries: chichest@dm.s.uinst.ac.uk		
COMMENT	BBRC/Dundee/Nottingham/Sanger/Sheffield/DMIST Gallus gallus cDNA sequencing project. This sequence is from the BBRC/Dundee/Nottingham/Sanger/Sheffield/DMIST cDNA collection, from a library constructed by Elizabeth Bosch. cDNA was prepared from RNA extracted from muscle, normalised, and poly A-tailed. EcoRI-NotI cut cDNA was then ligated into the vector. Vector: pUCscript II KS(+), Site_1: EcoRI, Site_2: NotI Host: Escherichia coli DH10B.		
FEATURES			
source	Location/Qualifiers 1..1522 /organism="Gallus gallus" /mol_type="mRNA" /strain="Layer and broiler" /db_xref="taxon:9031" /clone="CHEST189114" /clone_lib="CEORBN11" /dev_stage="adult"		
ORIGIN			
Alignment Scores:			
Pred. No.:	1,24e-82	Length:	1522
Score:	939.00	Matches:	176
Percent Similarity:	81.06%	Conservative:	38
Best local Similarity:	66.67%	Mismatches:	50
Query Match:	67.80%	Indels:	0
DB:	5	Gaps:	0

Dd		647	CTTTACTCTCCGAGGTCACTGCATCTTGGACAGTGAACAAGAAGTTTCAGATTCTT	:::     :::     :::     :::     :::	706
Oy		41	LeuIeuGIyLysSerLeuAspIleAlaProGIyAaspMeIleuAspLysValAlaArgArg	:::     :::     :::     :::     :::	60
Dd		707	CTGCTTGGAACAAGTCATAGATATATAGSACAAGGTGACATGCGTAGTAAGGTAGCAAGAG	:::     :::     :::     :::     :::	766
Oy		61	LeuSerLeuIleLysHisPProGIuCySerThrMetSerGIyGLyValAllegIunHis	:::     :::     :::     :::     :::	80
Dd		767	CTCTCTTTAGTAAAGCACCCGAGAGGCCAACGGCATGGCTGGGGGGAGAAAGCATAAGACAC	:::     :::     :::     :::     :::	826
Oy		81	LeuAlaLysGIuInGIyAAsnArgPheHisAspHeAspRIeLysPProProlenHisIleAlaLys	:::     :::     :::     :::     :::	100
Dd		827	CTGGGTCAAACCGAGACTGGSCAACAGTACCTTTCACACTTCCCATCACAACATGATGT	:::     :::     :::     :::     :::	886
Oy		101	AasnCyAasPheSerPheThrGIyLeuGIunHisValThraAspLysIleIleMetLysLys	:::     :::     :::     :::     :::	120
Dd		887	AACGTGATTTTTCTTTCTCCGACCTTCAAGACCCTTGCCAAACAAAGCCATCTTCAGAAA	:::     :::     :::     :::     :::	946
Oy		121	GIuLySGIuGIuGIyLIleGIuLysGIuInGIleLeuSerSerAlaIleAspIleAlaAla	:::     :::     :::     :::     :::	140
Dd		947	GAAAAAGAAAGAGATTCACAAGAAAGGGAATCCGTCTCTCCGTTAAAGACATCCCTCT	:::     :::     :::     :::     :::	1006
Oy		141	ThrValGIunHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe	:::     :::     :::     :::     :::	160
Dd		1007	GCTGCACAGCAGCTAGTGCTGCTCATATATTCACAGCGGACACACCGAGCCATGCTTTC	:::     :::     :::     :::     :::	1066
Oy		161	CysLysGIuArgAspLeuLeuPProGIuAsnAsnAlaValLeuValAlaSerGIyGLyVal	:::     :::     :::     :::     :::	180
Dd		1067	TGCATGAAAACACAGCATTTATTCACAAAACCTGCACCTGGTTGTATTCAGAGGAGT	:::     :::     :::     :::     :::	1126
Oy		181	AlaSerAsnPheTYrIleArgArgAlaLeuGIuIleLeuThrAsnAlaThrGlncYSerThr	:::     :::     :::     :::     :::	200
Dd		1127	GCAAGTAAATCAGTATATCAGAAAGACATGCAGACTCTGGCAAAAGCAACGGTTTTGCT	:::     :::     :::     :::     :::	1186
Oy		201	LeuLeuCySProPProArgLeuCYSerThraAsnGIyLIleMetIleAlaTPRasnGIY	:::     :::     :::     :::     :::	220
Dd		1187	TTTTCTGTCTCTCTCCCAAGCGTGTGCACCCATATATGATGTATGATTCATGCAGATGGC	:::     :::     :::     :::     :::	1246
Oy		221	IleGIuArgLeuArgAlaGIyLeuGIyLIleLeuHisAspRIleGIuGIyLIleArgTYrGIu	:::     :::     :::     :::     :::	240
Dd		1247	ATTGAAAGGTGGCGTGCAGAGATGTGTATTTATATCACTAGTACGATGGCATGCCGCTACAA	:::     :::     :::     :::     :::	1306
Oy		241	ProlYsCySPROLeuGIyValAspRIleSerLYSGluValGIyGIuAlaSerIleLysVal	:::     :::     :::     :::     :::	260
Dd		1307	CCAAAAGCTCCCTCGAATGATGATTTCCAAAAGAGTTGAAGAGGATTCATCAAAGTG	:::     :::     :::     :::     :::	1366
Oy		261	ProGIuLeuLys 264	:::     :::     :::     :::     :::	
Dd		1367	CCAAAGACTTAAG 1378	:::     :::     :::     :::     :::	

Search completed: November 10, 2005, 22:36:49  
Job time : 5276.16 secs

US-10-649-273-2_COPY_148_414 (1-267) x BX9349991 (1-1522)	
Oy	1 MetGjuaiahaiaaleumThriaalegleumThraasulyaVajgIupheP-ophelenua1 20
Db	587 ATGGAGGCTCAGGCACTTATCCATCAGACGACAGCAAGCAATATTAATTTCCCTCTTAGTT 646
Oy	21 LeumenuieserGjyGlyyhiisCysLeumenuaiaaleuValaIcIngllyuaIseraPphenenu 40

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus.p2n model

Run on: November 10, 2005, 16:59:00 ; Search time 886.614 Seconds  
(without alignments)  
2490.420 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414  
Perfect score: 1385  
Sequence: 1 MEAHALTRITRNKVFPPPLV.....DISKEVGASIKVPLQKMEI 267

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 9794790 seqs, 413490567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Command line parameters:

-MODEL=frame+ p2n.model -DEV=xld  
-O=/cgsm2\_1/USPTO.spool\_p/US10649273/runat\_02112005\_091339\_15609/app.query.fasta\_1.1429  
-DB=published.Applications\_NA -OPMT=fastcap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPEXT=0 -UNITS=bases -START=1 -END=1 -MATRIX=bloum62  
-TRNMS-human40.cdt -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10649273 @CGN 1.1 1034 @runat\_02112005\_091339\_15609  
-NCPUE=6 -ICPU=3 -NO\_MMAP -LARGEMEMORY -NEG\_SCORES=0 -WAIT -DISPLOCK=100  
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

Published Applications NA: \*  
1: /cgsm2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq: \*  
2: /cgsm2\_6/ptodata/1/pubpna/FCI\_NEW\_PUB.seq: \*  
3: /cgsm2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq: \*  
4: /cgsm2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq: \*  
5: /cgsm2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq: \*  
6: /cgsm2\_6/ptodata/1/pubpna/FCITIS\_PUBCOMB.seq: \*  
7: /cgsm2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq: \*  
8: /cgsm2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq: \*  
9: /cgsm2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq: \*  
10: /cgsm2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq: \*  
11: /cgsm2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq: \*  
12: /cgsm2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq: \*  
13: /cgsm2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq: \*  
14: /cgsm2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq: \*  
15: /cgsm2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq: \*  
16: /cgsm2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq: \*  
17: /cgsm2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq: \*  
18: /cgsm2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq: \*  
19: /cgsm2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq: \*  
20: /cgsm2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq: \*  
21: /cgsm2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq: \*  
22: /cgsm2\_6/ptodata/1/pubpna/US10J\_PUBCOMB.seq: \*  
23: /cgsm2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq: \*  
24: /cgsm2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq: \*  
25: /cgsm2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq: \*  
26: /cgsm2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq: \*  
27: /cgsm2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq: \*  
28: /cgsm2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1385	100.0	1416	18	US-10-120-988-177	Sequence 177, App
2	1385	100.0	1526	15	US-10-067-443-23	Sequence 23, Appl
3	1385	100.0	1526	19	US-10-649-273-23	Sequence 23, Appl
4	1385	100.0	1526	19	US-10-651-722-23	Sequence 23, Appl
5	1385	100.0	2197	15	US-10-067-443-21	Sequence 1, Appl
6	1385	100.0	2197	19	US-10-649-273-1	Sequence 1, Appl
7	1385	100.0	2572	19	US-10-651-722-1	Sequence 1, Appl
8	1385	100.0	2572	22	US-10-480-988-36	Sequence 36, Appl
9	1358.5	98.1	1387	15	US-10-067-443-21	Sequence 21, Appl
10	1358.5	98.1	1387	19	US-10-649-273-21	Sequence 21, Appl
11	1358.5	98.1	1387	19	US-10-651-722-21	Sequence 21, Appl
12	1358.5	98.1	1387	22	US-10-887-553A-1047	Sequence 1047, Ap
13	1348	97.3	1245	15	US-10-012-140-6	Sequence 6, Appl
14	1348	97.3	1245	15	US-10-012-140-4	Sequence 4, Appl
15	1304	86.9	1820	18	US-10-094-749-400	Sequence 400, App
16	1204	86.9	2890	21	US-10-723-860-7447	Sequence 7447, Ap
17	824.5	59.5	14364	15	US-10-067-443-20	Sequence 20, Appl
18	824.5	59.5	14364	19	US-10-649-273-20	Sequence 20, Appl
19	824.5	59.5	14364	19	US-10-651-722-20	Sequence 20, Appl
20	599	43.2	2734	24	US-10-450-763-20426	Sequence 20426, A
21	468	33.8	371	20	US-10-430-201-3118	Sequence 3118, Ap
22	468	33.8	371	20	US-10-430-201-3119	Sequence 3119, Ap
23	422.5	30.5	1917	19	US-10-424-599-66417	Sequence 66417, A
24	415.5	30.0	1601	26	US-11-097-143-34190	Sequence 34190, A
25	415.5	30.0	3656	26	US-11-097-143-34189	Sequence 34189, A
26	359	25.9	1628	20	US-10-437-963-11249	Sequence 11249, A
27	342	24.7	756	15	US-10-081-051-8	Sequence 8, Appl
28	342	24.7	4360	18	US-10-081-051-2	Sequence 2, Appl
29	321.5	23.2	1146	18	US-10-282-122A-14674	Sequence 14674, A
30	314.5	22.7	936	18	US-10-282-122A-8315	Sequence 8315, Ap
31	308.5	22.3	1044	18	US-10-282-122A-26972	Sequence 26972, A
32	308.5	22.3	94750	19	US-10-672-787-38	Sequence 38, Appl
33	304.5	22.0	1032	18	US-10-282-122A-11043	Sequence 11043, A
34	301.5	21.8	1000	20	US-10-343-561-50	Sequence 50, Appl
35	291.5	21.0	1026	9	US-09-815-242-7701	Sequence 7701, Ap
36	291.5	21.0	1026	18	US-10-282-122A-30016	Sequence 30016, A
37	291.5	21.0	1026	24	US-10-958-216-49	Sequence 49, Appl
38	289.5	20.9	1026	24	US-10-958-216-51	Sequence 51, Appl
39	289.5	20.9	1029	9	US-09-815-242-6946	Sequence 6946, Ap
40	289.5	20.9	1029	18	US-10-282-122A-22020	Sequence 22020, A
41	289.5	20.9	1830121	18	US-10-329-670-1	Sequence 1, Appl
42	289.5	20.9	1830121	21	US-10-158-865-1	Sequence 1, Appl
43	289.5	20.9	1830121	24	US-10-961-687-1	Sequence 1, Appl
44	285	20.6	372	21	US-10-357-990-61088	Sequence 61088, A
45	282.5	20.4	9967	24	US-10-795-159-529	Sequence 529, App

## ALIGNMENTS

RESULT 1  
US-10-120-988-177  
; Sequence 177, Application US/10120988  
; Public Information No. US20030219745A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Ren, Feiyun  
; APPLICANT: Wang, Duntui  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20030219745A1 Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 802CON



```

; CURRENT APPLICATION NUMBER: US/10/120,988
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 09/774,528
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pc_flt_genes Version 2.0
; SEQ ID NO 177
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205)..(1305)
; US-10-120-988-177

Alignment Scores:
Pred. No.: 1.9e-169      Length: 1416
Score: 1385.00          Matches: 267
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 18                  Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-120-988-177 (1-1416)
Qy      1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
Db      502 ATGAGGCTCATGACCTTACTATTAGGTGACCAATAAGTGAATTTCTTTTATGTT 561
Qy      21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db      562 CTTTGTATTTCTGAGGTCACGTCTGTGGCATTAAGTTCAGAGTTTCAGATTTCG 621
Qy      41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db      622 CTTCTTGAAAGTCTTTGGACATGACACAGGTGACAGCTTGACAGGTGGCAAGAGA 681
Qy      61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysValAlaIleGluHis 80
Db      682 CTTTCTTTAATAAATCAATCCAGAGTCTCCACATGAGTGGGGAAGGCCATGAACAT 741
Qy      81 LeuAlaLysGlnGlyAsnArgPheHisAspLeuIleLysProProLeuHisHisAlaLys 100
Db      742 TTGGCCAAACAGAGAAATAGATTTCATTTTGACATCAACCTCCCTTGATCATGCTAA 801
Qy      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db      802 AATTGATATTTTCTTTTACTGACCTTCACACGTTACTGATTAATAATAATGAAGAAG 861
Qy      121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
Db      862 GAAATAAGAGGAAAGTATTTGAGAAAGGGGCAAAATCCGTCTTCAGCAGCAGCAATTCCTGCC 921
Qy      141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
Db      922 ACAGTACAGCACACAAATGGCATGTATCTTGGAAGAAACACATCGGGCTATTCTGTT 981
Qy      161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyLysVal 180
Db      982 TGTAAAGAGAGAGACTTGTATCTCAATAATATGACAGTACTGTGCATCTGTGTGTGTC 1041
Qy      181 AlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaIleArgIleCysThr 200
Db      1042 GCAAGTAACTTTCTATATCCGAGAGCTCTGGAAATTTTAAACAAAGCACAACAGTGCAT 1101
Qy      201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleArgAsnGly 220
Db      1102 TTGTTGTGTCCTCTCCAGACTATGCACTGATTAATGGCATTAATGATTCAGAGAAATGCT 1161
Qy      221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgGlyArgLys 240
Db      1162 ATTGAAGACTACGTGCTGGCTTGGGCAATTTTACATGACATAGAAAGGCAATCCGCTATGAA 1221
```

```

Qy      241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260
Db      1222 CCAAAATGTCCTCTTGAGTAGACATATCAAAAGAAAGTGGAGAGCTTCATAAAGTA 1281
Qy      261 ProGlnLeuLysMetGluIle 267
Db      1282 CCACAAATTAATAATGAGATA 1302

RESULT 2
US-10-067-443-23
; Sequence 23, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-067-443-23

Alignment Scores:
Pred. No.: 2.13e-169      Length: 1526
Score: 1385.00          Matches: 267
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 15                  Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-23 (1-1526)
Qy      1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
Db      1 ATGAGGCTCATGACCTTACTATTAGGTGACCAATAAGTGAATTTCTTTTATGTT 60
Qy      21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db      61 CTTTGTATTTCTGAGGTCACGTCTGTGGCATTAAGTTCAGAGATTTCAGATTTTCGTG 120
Qy      41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db      121 CTTCTTGAAAGTCTTTGGACATGACACAGGTGACATGCTTTCACAAAGTGGCAAGAGA 180
Qy      61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysValAlaIleGluHis 80
Db      181 CTTTCTTTAATAAATCAATCCAGAGTCTCCACATGATAGGTGGAAAGCCATAGAACAT 240
Qy      81 LeuAlaLysGlnGlyAsnArgPheHisAspLeuIleLysProProLeuHisHisAlaLys 100
Db      241 TTGGCCAAACAGAGAAATGATTTTCATTTTGACATCAAACTCCCTTGATCATGCTAA 300
Qy      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db      301 AATTGATATTTTCTTTTACTGACTTCACACGTTTATGATTAATAATATGAAGAAAG 360
Qy      121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
Db      361 GAAAAAGAGAGAGTATTGAGAAAGGGCAAAATCTGTCTTCAACAGAGACATTTGCTGCC 420
Qy      141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
Db      421 ACAGTACAGCACACATGCGATGTCACTTGGAAGAAACACATCGGGCTATTCGTGTT 480
```



QY 161 CyslysglnargaspLeuProGlnAsnAsnAlaValIleuValAlaSerGlyGlyVal 180  
DB 481 TGTAAAGAGAGAGACTGTTACTTCAAAATATAGACATGCTGGCATCTGGTGCTGTC 540  
QY 181 AlSerAsnPhetYrIleArgArAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
DB 541 GCAGTAACCTTCTATATCCGACAGCTCTGGAAATTTTAAACAAACGCAACACAGTGCCT 600  
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
DB 601 TTGTTGTCTCTCTCCAGACTATGACATGATATGACATGATGATGATGATGATGATGAT 660  
QY 221 IleguValLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgGlyGlu 240  
DB 661 ATTGAAAGACTACGCTGCTGGCATTTTACATGACATAGAGGATCCGCTATGAA 720  
QY 241 ProlyCysProLeuGlyValAspIleSerIysGlyValIleGluAlaSerIleIysVal 260  
DB 721 CCAAAATGCTCTTGGAGTACATATCAAAAGAGTTGAGAGCTTCCATAAAGTA 780  
QY 261 ProGlnLeuIysMetGluIle 267  
DB 781 CCACATTTAAATAATGAGATA 801  
RESULT 3  
US-10-649-273-23  
; Sequence 23, Application US/10649273  
; Publication No. US20040043407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649,273  
; PRIOR FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 23  
; LENGTH: 1526  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-649-273-23  
Alignment Scores:  
Pred. No.: 2,136-169 Length: 1526  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 19 Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-649-273-23 (1-1526)  
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIysValGluPheProPheLeuVal 20  
DB 1 ATGAGGCTCATGCACTTACTATTAGTTGACCAATAAAGTAGAATTTCTTTTACTT 60  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValIleSerAspPheLeu 40  
DB 61 CTTTGAATTTCTGAGGTCACGTCGCTGTTGCAATTAGTCAAGAGTTCAGATTTCTCG 120  
QY 41 LeuLeuGlyIysSerLeuAspIleAlaIleProGlyAspMetLeuAspIysValAlaArgArg 60  
DB 121 CTTTGTGAAAGGCTTTTGACATAGCACAGGTGACATGCTTGACAAAGTGCAAGAA 180  
QY 61 LeuSerLeuIleIysHisProGluCysSerThrMetSerGlyGlyIysValIleGluHis 80  
DB 181 CTTTCTTAAATAAATCATCCAGAGTCTCCACATGATGCTGGAGAAAGCCATAGAACT 240

QY 81 LeuAlaIysGlnGlyAsnArgPheHisPheAspIleIysProProLeuHisHisAlaIys 100  
DB 241 TTGGCCAAACAAAGAAATGATATTTCATTTTGCATCAAACTCCCTTGATCATGCTAA 300  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValIleThrAspIysIleIleMetIysIys 120  
DB 301 AATTGATTTTCTTTTACTGACTTCAACAGTTACTGATTAATAATTAATGAAGAAAG 360  
QY 121 GluIysGluGluGlyIleGluIysGlyGlnIleLeuSerSerAlaIleAspIleAlaIle 140  
DB 361 GAAAAAGGAAAGCATTTTGAAGAGGGCAAACTCTCTTACGACGACAGCATGCTGCTCC 420  
QY 141 ThrValGlnHisThrMetAlaCysHisIleuValIysArgThrHisArgAlaIleLeuPhe 160  
DB 421 ACAGTACAGCACACATGCGATGTCATCTTGGAAGAAACATCATGGGCTATTCTGTTT 480  
QY 161 CysIysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyGlyVal 180  
DB 481 TGTAAAGCAGAGACTGTTTACCTCAAAATATGACGTCTGTTGCATCTGCTGCTGTC 540  
QY 181 AlSerAsnPhetYrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
DB 541 GCAGTAACCTTCTATATCCGACAGCTCTGGAAATTTTAAACAAACGCAACACAGTGCCT 600  
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
DB 601 TTGTTGTCTCTCTCCAGACTATGACATGATATGATGATGATGATGATGATGATGAT 660  
QY 221 IleguValLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgGlyGlu 240  
DB 661 ATTGAAAGACTACGCTGCTGGCATTTTACATGACATAGAGGATCCGCTATGAA 720  
QY 241 ProlyCysProLeuGlyValAspIleSerIysGlyValIleGluAlaSerIleIysVal 260  
DB 721 CCAAAATGCTCTTGGAGTACATATCAAAAGAGTTGAGAGCTTCCATAAAGTA 780  
QY 261 ProGlnLeuIysMetGluIle 267  
DB 781 CCACATTTAAATAATGAGATA 801  
RESULT 4  
US-10-651-722-23  
; Sequence 23, Application US/10651722  
; Publication No. US20040048302A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 DIV  
; CURRENT APPLICATION NUMBER: US/10/651,722  
; PRIOR FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 23  
; LENGTH: 1526  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-651-722-23  
Alignment Scores:  
Pred. No.: 2,136-169 Length: 1526  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 19 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-651-722-23 (1-1526)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
DB 1 ATGAGAGCTCATGCACTTACTTATAGTTGACCAATTAAGTAATTTCTTTTATGTT 60  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 61 CTTTGAATTTCTGGAGGTCATGCTGCTGTTGGCATTAAGTTCAAGAGTTTCAGATTTTCG 120  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
DB 121 CTTCTTGGAAAGCTTTGGACATAGCACAGGTGACATGCTTGACCAAGGTGGCAAGAA 180  
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
DB 181 CTTTCTTTAATAACATCCAGAGTCTCCACCATGATGATGGTGGAAAGCCATTAACAT 240  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
DB 241 TTGGCCAAACAGAGAAATGATTTTCATTTTGACATCAAACTCCCTTGACATCATCTAA 300  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
DB 301 AATTGTGATTTTCTTTTACTGACCTTCACACGTTTACTGATTAATAATATGAAAAAG 360  
QY 121 GluLysGlnGlyGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140  
DB 361 GAAAAAGAGAGATTTGAGAGAGGCAAACTGCTTCACGACGACATTCCTGCC 420  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
DB 421 ACAGTACACGACACATGCACTGTCATCTTGGAAGAAACACACATCGGCTATTTCTGTT 480  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
DB 481 TGTAAAGACAGAGACTGTGTACTCAAAATATGACATGCTGTGCTGCTGCTGCTG 540  
QY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
DB 541 GCAAGTAACTTCTATATCCGACAGACTCTGGAATTTTAAACAAAGCAACACAGTGCCT 600  
QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrPasnGly 220  
DB 601 TTGTTGTCTCTCTCCACGACTATGCACTGATTAATGCACTTATGATTCATGAAATGT 660  
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 240  
DB 661 ATTGAAGACTACGTCGTGGCTTGGCATTTTACATGACATGAAGGCAATCCGCTATGAA 720  
QY 241 ProLysCysProLeuGlyValAspIleSerLysGlnValGlyGluAlaSerIleLysVal 260  
DB 721 CCAAAATGTCCTCTTGGAGTAGACATATCAAAAGAAAGTTGGAGAGCTTCCATAAAAGTA 780  
QY 261 ProGlnLeuLysMetGluIle 267  
DB 781 CCAACATTAATAATGAGATA 801

RESULT 5  
US-10-067-443-1  
; Sequence 1, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10

; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 2197  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (231) .. (1472)  
US-10-067-443-1

Alignment Scores:  
Pred. No.: 3,678-169 Length: 2197  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 15 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-067-443-1 (1-2197)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
DB 672 ATGAGAGCTCATGCACTTACTTATAGTTGACCAATTAAGTAATTTCTTTTATGTT 731  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 732 CTTTGAATTTCTGGAGGTCATGCTGCTGTTGGCATTAAGTTCAAGAGTTTCAATTTTCTG 791  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
DB 792 CTTCTTGGAAAGCTTTGGACATAGCACAGGTGACATCTTGACCAAGGTGGCAAGAA 851  
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
DB 852 CTTTCTTTAATAACATCCAGAGTCTCCACCATGATGATGGAGAAACATTAAGAAAT 911  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
DB 912 TTGGCCAAACAGAGAAATGATTTTCATTTTGACATCAAACTCCCTTGACATCATCTAA 971  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
DB 972 AATTGTGATTTTCTTTTACTGCACTTCAACGTTTACTGATTAATAATGAAAAAG 1031  
QY 121 GluLysGlnGlyGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140  
DB 1032 GAAAAAGAGAGATTTGAGAGAGGCAAACTGCTTCACGACACACATTCGCTGCC 1091  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
DB 1092 ACAGTACACGACACATGCACTGTCATCTTGGAAGAAACACATCGGCTATTTCTGTT 1151  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
DB 1152 TGTAAAGACAGAGACTGTGTACTCAAAATATGCACTGCTGTCATCTGCTGCTG 1211  
QY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
DB 1212 GCAAGTAACTTCTATATCCGACAGACTCTGGAATTTTAAACAAACCAACACAGTGCCT 1271  
QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrPasnGly 220  
DB 1272 TTGTTGTCTCTCTCCACGACTATGCACTGATTAATGCACTTATGATTCATGAAATGT 1331  
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 240  
DB 1332 ATTGAAGACTACGTCGTGGCTTGGCATTTTACATGACATTAAGAGCAATCCGCTATGAA 1391  
QY 241 ProLysCysProLeuGlyValAspIleSerLysGlnValGlyGluAlaSerIleLysVal 260  
DB 1392 CCAAAATGTCCTCTTGGAGTAGACATATCAAAAGAAAGTTGGAGAGCTTCCATAAAAGTA 1451

QY 261 ProGlnLeuLysMetGluile 267  
 Db 1452 CCACAATTAAAAATGAGATA 1472

RESULT 6

US-10-649-273-1  
 ; Sequence 1, Application US/10649273  
 ; Publication No. US20040043407A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bristol-Myers Squibb Company  
 ; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
 ; FILE REFERENCE: D0073 CNT  
 ; CURRENT APPLICATION NUMBER: US/10/649,273  
 ; CURRENT FILING DATE: 2003-08-27  
 ; PRIOR APPLICATION NUMBER: US 60/266,518  
 ; PRIOR FILING DATE: 2001-02-05  
 ; PRIOR APPLICATION NUMBER: US 10/067,443  
 ; PRIOR FILING DATE: 2002-02-05  
 ; PRIOR APPLICATION NUMBER: US 60/282,814  
 ; PRIOR FILING DATE: 2001-04-10  
 ; NUMBER OF SEQ ID NOS: 71  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 1  
 ; LENGTH: 2197  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (231)..(1472)  
 US-10-649-273-1

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
3.67e-169	1385.00	2197	267
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	19	Gaps:	0

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-649-273-1 (1-2197)

QY 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
 Db 672 ATGGAGGCTCATGCACTTACTATTAGGTGACCAATAAGATTAATTTCTTTTATGTT 731  
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
 Db 732 CTTTGAATTTCTGGAGGTCACTGTCTGTTGGCATTAGTTCAAGAGTTTCAGATTTTCTG 791  
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 60  
 Db 792 CTTCTTGGAAAGCTTTTGACATACGACAGGTGACATCTTTCACAGGTGGCAAGAGA 851  
 QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
 Db 852 CTTTCTTTAATAAATCATCCAGAGTGTCTCCACCATGATGTTGGTGAAGCCATGAACAT 911  
 QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProPheLeuHisAlaLys 100  
 Db 912 TTGGCCAAACAAAGCAATATGATTCAATTGACATCAAACTCCCTTGCATCATGCTAAA 971  
 QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
 Db 972 AATTGATTTTCTTTTACTGACCTTCAACAGTTACTGATTAATAAATATATATAAAG 1031  
 QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140  
 Db 1032 GAAAAAGAGAGAGATTAAGAAAGGGCAATCTGTCTTACAGACAGACATTTCTGCC 1091  
 QY 144 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
 Db 1092 ACAGTACAGACACAAATGGATGTCACTTGTGAAAGAACACATCGGGCATATCTGTTT 1151

QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
 Db 1152 TGTAAAGCAGAGACTTGTACTCTCAAAATAATGAGTACTGTTCATCTGTGTGTCTC 1211  
 QY 161 AlaSerAspPheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
 Db 1212 GCAAGTAACTTTATATCCGCAAGCTCTGAAATTTTAAACAAACCAACACAGTCACT 1271  
 QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
 Db 1272 TTGTTGTGTCCCTCCCTCCAGACTATGACATGATTAATGGCATTAATGATGATGATG 1331  
 QY 221 IleGluValLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgGlyVal 240  
 Db 1332 ATTTGAAGACTACGCTGGCTGGCTTGGGCAATTTTACATGATAGAAAGCATCCGCTATGAA 1391  
 QY 241 ProLysCysProLeuGlyValAlaAspIleSerLysGluValGlyGluAlaSerIleLysVal 260  
 Db 1392 CCAAAATGCTCTTGTGAGTACATATCAAAAGAGTTGAGAAAGCTTCCATAAAGTA 1451  
 QY 261 ProGlnLeuLysMetGluile 267  
 Db 1452 CCACAATTAAAAATGAGATA 1472

RESULT 7

US-10-651-722-1  
 ; Sequence 1, Application US/10651722  
 ; Publication No. US20040048302A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bristol-Myers Squibb Company  
 ; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
 ; FILE REFERENCE: D0073 DIV  
 ; CURRENT APPLICATION NUMBER: US/10/651,722  
 ; CURRENT FILING DATE: 2003-08-29  
 ; PRIOR APPLICATION NUMBER: US 60/266,518  
 ; PRIOR FILING DATE: 2001-02-05  
 ; PRIOR APPLICATION NUMBER: US 10/067,443  
 ; PRIOR FILING DATE: 2002-02-05  
 ; PRIOR APPLICATION NUMBER: US 60/282,814  
 ; PRIOR FILING DATE: 2001-04-10  
 ; NUMBER OF SEQ ID NOS: 71  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 1  
 ; LENGTH: 2197  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (231)..(1472)  
 US-10-651-722-1

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
3.67e-169	1385.00	2197	267
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	19	Gaps:	0

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-651-722-1 (1-2197)

QY 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
 Db 672 ATGGAGGCTCATGCACTTACTATTAGGTGACCAATAAGATTAATTTCTTTTATGTT 731  
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
 Db 732 CTTTGAATTTCTGGAGGTCACTGTCTGTTGGCATTAGTTCAAGAGTTTCTG 791  
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 60  
 Db 792 CTTCTTGGAAAGCTTTTGACATACGAGTGAATGCTTGAACAGGTGGCAAGAGA 851

```
QY 61 LeuSerLeuIleYshISProGluCySerThrMetSerGlyValAlaIleGluHis 80
Db 852 CTTTCTTTAATAAACAATCCAGAGTGTCCACCATGATGGTGGGAAAGCCATAGAAAT 911
QY 81 LeuAlaIleGlnGlyValAsnArgPheHisPheAspIleYsProProLeuHisHisAlaIys 100
Db 912 TTGGCCAAACAGAGAAATGATTCATTTTGAACATCAAACTCCCTTCATCATGCTAA 971
QY 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspIleYsIleIleMetIysIys 120
Db 972 AATTGTGATTTTCTTTTACTGAGACTTCAACACGTTACTGATTAATAATATGAAAAAG 1031
QY 121 GluIysGlnGluGlyIleGluIysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
Db 1032 GAAAGAGGAGGATGATGAGAAAGGCAAAATCCTGCTTCGACGACAGACATTCCTGCTC 1091
QY 141 ThrValGlnHisThrMetAlaCyHisIleuValIysArgThrHisArgAlaIleLeuPhe 160
Db 1092 ACAGTACGACACACAAATGGCATGTCATCTTGGAAGAAACACATCGGGCTATTCGTTT 1151
QY 161 CysIysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyVal 180
Db 1152 TGTAAAGCAGAGACTTGTACCTCAAAATATGACAGTACTGGTTCATCTGGTGTGTC 1211
QY 181 AlaSerAsnPheYrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
Db 1212 GCAAGTAACTCTTATATCGCAGAGCTCTGGAAATTTTAAACAAACGCAACACAGTGCAT 1271
QY 201 LeuLeuCySPProProProArgLeuCySPThrAspAsnGlyIleMetIleAlaTrpAsnGly 220
Db 1272 TTGTTGTGTCTCTCCCTCCAGACTATGACATGCTATATGATGATGATGATGATGATG 1331
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTrpGlu 240
Db 1332 ATTAAGAACTACGAGCTGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1391
QY 241 ProIysCySPProLeuGlyValAspIleSerIysGlnValGlyGlnAlaSerIleYsVal 260
Db 1392 CCAAAATCTCTCTTGGAGTACATATCAAAAGAGAGTGGAGAGCTTCCATAAAGTA 1451
QY 261 ProGlnLeuIysMetGluIle 267
Db 1452 CCACAAATTAATAATGAGATA 1472

RESULT 8
US-10-480-988-36
; Sequence 36, Application US/10480988
; Publication No. US20050069877A1
; GENERAL INFORMATION:
; APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;
; APPLICANT: SARMAKAR, Anita; HAPALIA, April J.A.;
; APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;
; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;
; APPLICANT: HONCHELL, Cynthia D.; NGUYEN, Daniel B.;
; APPLICANT: LU, Duang Aina M.; LEE, Ernestine A.;
; APPLICANT: YUE, Henry; FORSYTHE, Ian J.;
; APPLICANT: BARROSO, Ines; RAMKOMAR, Jayalaxmi;
; APPLICANT: GRIFFIN, Jennifer A.; LI, Joana X.;
; APPLICANT: YANG, Junning; THANGAVELU, Kavitha;
; APPLICANT: GIETZEN, Kimberly J.; DING, Li;
; APPLICANT: BAUGHN, Mariah R.; BOROMSKY, Mark L.;
; APPLICANT: YAO, Monique G.; CHAMLA, Narinder K.;
; APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
; APPLICANT: LEE, Sally; BECHA, Shanya D.;
; APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;
; APPLICANT: ELIOTT, Vicki S.; LUO, Wen;
; APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;
; APPLICANT: LU, Yan; ZEBARADIAN, Yeganeh
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PF-1040 USN
; CURRENT APPLICATION NUMBER: US/10/480,988
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: PCT/US02/19360
```

```
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/300,508
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/303,445
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/305,405
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/311,442
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 60/314,821
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/315,992
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 60/378,205
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 36
; LENGTH: 2572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7632424CB1
; US-10-480-988-36

Alignment Scores:
Pred. No.: 4 65e-169 Length: 2572
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-480-988-36 (1-2572)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIysValGluPheProPheLeuVal 20
Db 588 ATGGAGGCTCATGACCTTATCTATGAGTTGACCAATTAAGTGAATTTCTTTTAGTT 644
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 645 CTTTGTGATTTCTGGAGGTCACTGCTGTTGGCATTTGTTCAAGAGATTCAGATTTTCTG 704
QY 41 LeuLeuGlyIysSerLeuAspIleAlaProGlyAspMetLeuAspIysValAlaArgArg 60
Db 705 CTTTCTTGAAGCTTTTGGACATGACACAGGATGACATGCTTGAACAGGTGGCAAGAGA 764
QY 61 LeuSerLeuIleYshISProGluCySerThrMetSerGlyIlyIysAlaIleGluHis 80
Db 765 CTTTCTTTAATAAACAATCCAGAGTGTCCACCATGATGGTGGGAAAGCCATAGAAAT 824
QY 81 LeuAlaIleGlnGlyValAsnArgPheHisPheAspIleYsProProLeuHisHisAlaIys 100
Db 825 TTGGCCAAACAGAGAAATGATTCATTTTGAACATCAAACTCCCTTCATCATGCTAA 884
QY 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspIleYsIleIleMetIysIys 120
Db 885 AATTGTGATTTTCTTTTACTGAGACTTCAACACGTTACTGATTAATAATATGAAAAAG 944
QY 121 GluIysGlnGluGlyIleGluIysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
Db 945 GAAAGAGGAGGATGATGAGAAAGGCAAAATCCTTTCGACGACAGACATTCGCTGCC 1004
QY 141 ThrValGlnHisThrMetAlaCyHisIleuValIysArgThrHisArgAlaIleLeuPhe 160
Db 1005 ACAGTACGACACACAAATGGCATGTCATCTTGGAAGAAACACATCGGGCTATTCGTTT 1064
QY 161 CysIysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyVal 180
Db 1065 TGTAAAGCAGAGACTTGTATCTCAAAATATGACAGTACTGGTTCATCTGGTGTGTC 1124
QY 181 AlaSerAsnPheYrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
```

```
Db 1125 GCAAGTAACTTCTATATCCGACAGCTCTGGAAATTTTAAACAAACCAACACAGTCCACT 1184
Qy 201 LeuLeuCySPProProArgLeuCySPThraSPasnGly1IleMet1IleAlaTPasnGly 220
Db 1185 TTGTGTGTCTCCCTCCGACGACTATGCACTGATTAATGCGATTATGATTCATCGAATGCT 1244
Qy 221 ILeGluAlaGLeuAlaGlyLeuGly1IleLeuHisAsp1IleGluGly1IleArgTyGlu 240
Db 1245 ATTGAAGAACTACGCTGGCTGGCTTGCGCATTTTACATGACATAGAAAGCATCCGCTATGAA 1304
Qy 241 ProLySPSPProLeuGlyValaSPleSPSerLySPGluValGlyGluAlaSer1IleLySPVal 260
Db 1305 CCAAAATGTCCTCTTGAGTAGACATATCAAAAGAAAGTTGAGAAAGCTTCATAAAGTA 1364
Qy 261 ProGluLeuLySPMetGluIle 267
Db 1365 CCACATTTAAATAATGAGATA 1385

RESULT 9
US-10-667-443-21
; Sequence 21, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-667-443-21

Alignment Scores:
Pred. No.: 5,276-166 Length: 1387
Score: 1358.50 Matches: 266
Percent Similarity: 91.10% Conservative: 0
Best Local Similarity: 91.10% Mismatches: 1
Query Match: 98.09% Indels: 25
Gaps: 1

US-10-649-273-2_COPY_148_414 (1-267) x US-10-667-443-21 (1-1387)
Qy 1 MetGluAlaHisAlaLeuThr1IleArgLeuThraSPLeuValGluPheProPheLeuVal 20
Db 465 ATGAGGCTCATGACCTTACTATATAGTGTGACCAATTAAGTGAATTCCTTTTATAGT 524
Qy 21 LeuLeu1IleSerGlyGlyHisCySPLeuLeuAlaLeuValGlnGlyValaSerAspPheLeu 40
Db 525 CTTTGATTTCTGAGGCTGACGCTGCTGTTGGCATTAAGTCAAGGAGTTTCAATTTTCTG 584
Qy 41 LeuLeuGlyLySPSerLeuAsp1IleAlaProGlyAspMetLeuAspLySPValaIlaArgArg 60
Db 585 CTTCTTGAAAGCTCTTTGACATAGCACCAAGGTGACATGCTTGACCAAGGTGACAAAGAA 644
Qy 61 LeuSerLeu1IleLySPHisAspProGluCySPSerThraMetSerGlyGlyLySPAla1IleGluHis 80
Db 645 CTTTCTTAATAAACAATCCAGAGTGTCCACCAATGAGTGTGGGAAAGCCATAGAGCAT 704
Qy 81 LeuAlaLySPGlnGlyAsnArgPheHisAspAsp1IleLySPProLeuHisHisAlaLySP 100
Db 705 TTGGCCAAACAAAGAAATGATTTTCATTTTGCATCAAAACCTCCCTGCATCATCTAA 764
Qy 101 AsnCySPAspPheSerPheThrGlyLeuGlnHisValThraSPLySP1IleLeuMetLySP 120
```

```
Db 765 AATTGGAATTTTCTTTACTGACTTCAACAGCTTAATGATAAATTAATGAAGAAAG 824
Qy 121 GluLySPGluGlyGly1Ile----- 126
Db 825 GAAAAAGAGAGAGTATATTTCTAATTAGTAAGTTGAACAGATTAATATCTCGATTG 884
Qy 127 -----GluLySPGlyGlu1IleLeuSerSerAla 135
Db 885 TGCGTAAATAATAGCTGCTCATTTCTGACAGTATAGAGAGGGCAAAATCCTGCTTACGA 944
Qy 136 AlaSP1IleAla1IleThrValGlnHisThraMetAlaCySPHisLeuValaIleGlyHis 155
Db 945 GCGACATTTGCTCCCAAGTACAGCACAAATGGCATGTTCATCTTGAAAGAACACAT 1004
Qy 156 ArgAla1IleLeuPheCySPLySPGlnArgAspLeuProGlnAsnAsnAlaValaLeuVal 175
Db 1005 CGGGCTATTTCTGTTTGTAAAGCAGAGACTGTGTACCTCAAAATTAATGACATGAGT 1064
Qy 176 AlaSerGlyGlyVala1IleSerAspPheTy1IleArgArgAlaLeuGlu1IleLeuThraSP 195
Db 1065 GCATCGTGTGTGTGCGCAAGTAACTTATATCCGACAGCTCTGGAAATTTTAAACAAAC 1124
Qy 196 AlaThrGlnCySPThraLeuCySPProProArgLeuCySPThraSPasnGly1IleMet 215
Db 1125 GCAACACAGTGCACCTTGTGTGCTCCTCCACAGCTATGACATGATATAGCATTAAG 1184
Qy 216 ILeAlaTPasnGly1IleGluArgLeuArgAlaGlyLeuGly1IleLeuHisAsp1IleGlu 235
Db 1185 ATTGATGGAATAGTATTAAGAACTACGTGTGGCTTGCGCATTTTACATGACATAGAA 1244
Qy 236 Gly1IleArgTyGluPProLySPSPProLeuGlyValaSP1IleSerLySPGluValaGlyGlu 255
Db 1245 GGCATCCGCTATGAACCAAAATGTCTCTTGAGTGAATGACATATCAAAAGAGTTGAGAA 1304
Qy 256 AlaSer1IleLySPValaProGluLeuLySPMetGluIle 267
Db 1305 GCTTCATTAAGTACCAATTAATAATGAGATA 1340

RESULT 10
US-10-649-273-21
; Sequence 21, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-649-273-21

Alignment Scores:
Pred. No.: 5,276-166 Length: 1387
Score: 1358.50 Matches: 266
Percent Similarity: 91.10% Conservative: 0
Best Local Similarity: 91.10% Mismatches: 1
Query Match: 98.09% Indels: 25
Gaps: 1

US-10-649-273-2_COPY_148_414 (1-267) x US-10-649-273-21 (1-1387)
```

```

QY      1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
Dh      465 ATGAGAGCTCATGACCTACTATTAGTGAACCAATTAAGTGAATTCCTTTTAGTT 524
QY      21 LeuLeuIleSerGlyValHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Dh      525 CTTTGAATTTCTGGAGGTCACGTCTGTGGCATTAGTCAAGAGATTTCAGATTTTCTG 584
QY      41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Dh      585 CTTCTTGAAAGTCTTTTGACATGACCAAGGTGACATGCTTGACAAAGGTGGCAAGAGA 644
QY      61 LeuSerLeuIleLysHisProGlyLysCysSerThrMetSerGlyLysAlaIleGluHis 80
Dh      645 CTTTCTTAATTAACATCCAGAGTCTCCACCATGAGTGGGGAAAGCCATAGAGCAT 704
QY      81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
Dh      705 TTGGCCAAACAAAGAAATAGATTTCATTGTCATCAAACTCCCTTGATCATGCTAA 764
QY      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Dh      765 AATTGTGATTTTCTTTTAACTGGACTTCAACAGTTACGATTAATAATTAATGAAGAA 824
QY      121 GluLysGlnGluGlyIle----- 126
Dh      825 GAAAAAGAGAAAGATATTTCTATTAGTAAGTTGAACAGATAAATATTCCTGATG 884
QY      127 -----GluLysGlyGlnIleLeuSerSerAla 135
Dh      885 TGCCCTAAAAATAGCTGCTCATTTCTGACAGGTATGAGAAAGGGCAAAATCCTGCTCAGCA 944
QY      136 AlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHis 155
Dh      945 GCAGCATTTGCTGCACAGTACAGCAACAATGGCATGTCTGTGTGAAGAAAGAACACAT 1004
QY      156 ArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuVal 175
Dh      1005 CGGGCTATTCTGTTTGTGAAGAGACATTTGTTACCTCAAAATTAATGACATACGTGTT 1064
QY      176 AlaSerGlyGlyValAlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsn 195
Dh      1065 GCATCTGGTGGTGCAGCAAGTAATCTTATATCCGAGAGCTCTGGAATTTTAAACAAC 1124
QY      196 AlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMet 215
Dh      1125 GCAACACAGTGCATTTGTGTGTCTCTCCAGACTATGCACTGATATGCGCATATG 1184
QY      216 IleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235
Dh      1185 AATGATGAGAAATGATTAAGAAAGACTACGTGGTGGCTTGAGCATTTTACATGACATAGA 1244
QY      236 GlyIleLeuGlyArgLysCysProLeuGlyValAspIleSerLysGlnValGlyGln 255
Dh      1245 GGCATCCGCTATGAACCAAAATGTCTCTTGAGATGACATATCAAAAGAAAGTTGAGAA 1304
QY      256 AlaSerIleLysValProGlnLeuLysMetGluIle 267
Dh      1305 GCTTCATTAATAAGTACCAATTAATAATGAGAGATA 1340

```

RESULT 11

```

US-10-651-722-21
; Sequence 21, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443

```

```

; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-651-722-21

Alignment Scores:
Pred. No.: 5,27e-166 Length: 1387
Score: 1358.50 Matches: 266
Percent Similarity: 91.10% Conservative: 0
Best Local Similarity: 91.10% Mismatches: 1
Query Match: 98.09% Indels: 25
DB: 19 Gaps: 1

US-10-649-273-2_COPY_148_414 (1-267) x US-10-651-722-21 (1-1387)
QY      1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
Dh      465 ATGAGAGCTCATGACCTACTATTAGTGAACCAATTAAGTGAATTCCTTTTAGTT 524
QY      21 LeuLeuIleSerGlyValHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Dh      525 CTTTGAATTTCTGGAGGTCACGTCTGTGGCATTAGTCAAGAGATTTCAGATTTTCTG 584
QY      41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Dh      585 CTTCTTGAAAGTCTTTTGACATGACCAAGGTGACATGCTTGACAAAGGTGGCAAGAGA 644
QY      61 LeuSerLeuIleLysHisProGlyLysCysSerThrMetSerGlyLysAlaIleGluHis 80
Dh      645 CTTTCTTAATTAACATCCAGAGTCTCCACCATGAGTGGGGAAAGCCATAGAGCAT 704
QY      81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
Dh      705 TTGGCCAAACAAAGAAATAGATTTCATTGTCATCAAACTCCCTTGATCATGCTAA 764
QY      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Dh      765 AATTGTGATTTTCTTTTAACTGGACTTCAACAGTTACGATTAATAATTAATGAAGAA 824
QY      121 GluLysGlnGluGlyIle----- 126
Dh      825 GAAAAAGAGAAAGTATATTTCTAATTAGTAAGTTGAACAGATAAATATTCCTGATG 884
QY      127 -----GluLysGlyGlnIleLeuSerSerAla 135
Dh      885 TGCCCTAAAAATAGCTGCTCATTTCTGACAGGTATGAGAAAGGGCAAAATCCTGCTTACGA 944
QY      136 AlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHis 155
Dh      945 GCAGCATTTGCTGCACAGTACAGCAACAATGGCATGTCTGTGTGAAGAAAGAACAT 1004
QY      156 ArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuVal 175
Dh      1005 CGGGCTATTCTGTTTGTGAAGAGAGACTTGTTCCTCAAAATTAATGACATGCTGTT 1064
QY      176 AlaSerGlyGlyValAlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsn 195
Dh      1065 GCATCTGGTGGTGCAGCAAGTAATCTTATATCCGAGAGCTCTGGAATTTTAAACAAC 1124
QY      196 AlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMet 215
Dh      1125 GCAACACAGTGCATTTGTGTGTCTCTCCAGACTATGCACTGATATGCGCATATG 1184
QY      216 IleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235
Dh      1185 AATGATGAGAAATGATTAAGAAAGACTACGTGGTGGCTTGAGCATTTTACATGACATAGA 1244

```



```
QY 236 GYIIEARGYRGUProlyseCysProleuGlyValAspIleSerIysGluValGlyGlu 255
DB 1245 GGATCGCGCTATGAAACCAAAATGTCCTTGAGTAGCATATCAAAAAGTTGGAGAA 1304
QY 256 AAlserIleIysValProGlnLeuIysMeGluIle 267
DB 1305 GCTTCCATTAAGTACCAATTTAAATAATGGAGATA 1340

RESULT 12
US-10-887-553A-1047
; Sequence 1047, Application US/10887553A
; Publication No. US20050085436A1
; GENERAL INFORMATION:
; APPLICANT: Garza, Dan
; APPLICANT: Li, Hao
; TITLE OF INVENTION: Method to treat conditions associated
; TITLE OF INVENTION: with insulin signalling dysregulation
; FILE REFERENCE: 4-33262
; CURRENT APPLICATION NUMBER: US/10/887, 553A
; PRIOR FILING DATE: 2004-07-08
; PRIOR APPLICATION NUMBER: 60/485,883
; PRIOR FILING DATE: 2003-08-07
; NUMBER OF SEQ ID NOS: 1208
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1047
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: human
US-10-887-553A-1047

Alignment Scores:
Pred. No.: 5,27e-166 Length: 1387
Score: 1358.50 Matches: 266
Percent Similarity: 91.10% Conservative: 0
Best Local Similarity: 91.10% Mismatches: 1
Query Match: 98.09% Indels: 25
DB: Gaps: 1

US-10-649-273-2_COPY_148_414 (1-267) x US-10-887-553A-1047 (1-1387)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIysValGluPheProPheLeuVal 20
DB 465 ATGAGAGCTCATGACCTTACTATTAGGTGACCAATTAAGATGATTTCTTTTAAAGTT 524
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 525 CTTTGTATTTCTGGAGGTCACGTGCTGTGTGTCATTAGTTCAGAGATTTCGATTTCGTG 584
QY 41 LeuLeuGlyIysSerLeuAspIleAlaProGlyAspMetLeuAspIysValAlaArgArg 60
DB 585 CTTCTTGGAAGAGCTTTTGACATAGACACAGGTGACATGCTTGACAAAGTGGCAAGAA 644
QY 61 LeuSerLeuIleIysHisAspGluCysSerThrMetSerGlyGlyIysAlaIleGluHis 80
DB 645 CTTTCTTTAATAAACAATCCAGAGTGTCCACCATGAGTGGGAAAGCCATAGAGCAT 704
QY 81 LeuAlaIysGlnGlyAsnArgPheHisPheAspIleIysProProLeuHisHisAlaIys 100
DB 705 TTGGCCCAACAAAGAAATAGATTTCATTTCATTCACAAACCTCCCTGCATCATGCTTAA 764
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIysIleIleMetIysIys 120
DB 765 AATTGTGATTTTCTTTTACTGACCTTCACACGTTACTGATTAATAATATGAAAAAG 824
QY 121 GluIysGluGluGlyIle----- 126
DB 825 GAAAAAAGAGAGCTATATTCTAATTAGTAAGTGAACAGATAAATATTCCTGAGATTG 884
QY 127 -----GluIysGlyGlnIleLeuSerSerAla 135
DB 885 TGCCTAAAAATAGCTGCTATTTCGACAGATAGAGAAAGGGCAAAATCTCTGCTTACACA 944
QY 136 AlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisIleValIysArgThrHis 155
```

```
DB 945 GCAGACATTGCTCCACAGTACAGACACAAATGGCATGTCTGTGAAAAAGAACAT 1004
QY 156 ArgAlaIleLeuPheCysIysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleVal 175
DB 1005 CCGGCTATTCTCTTTGTTAGAGAGAGAGACTTGTTACCTCAAAATATATGACTACTGCTT 1064
QY 176 AAlserGlyIysValAlaSerAsnAspPheIleArgArgAlaLeuGluIleLeuThrAsn 195
DB 1065 GCATCTGGTGTGTGCGAAAGTAACTTCTATATCCGACAGAGCTCTGGAAATTTTAACAAC 1124
QY 196 AlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMet 215
DB 1125 GCACACAGTGCATTTGTGTGTCCTCTCCACACATATGACATGATATATGCAATTATG 1184
QY 216 IleAlaTPAsnGlyIleGluArgLeuArgAlaGlyIleGluIleLeuHisAspIleGlu 235
DB 1185 ATTGATGAGATGATGATTGAAAGACTACGTGTGGCTTGGCATTTTACATATACATAGAA 1244
QY 236 GlyIleArgTYRGUProlyseCysProleuGlyValAspIleSerIysGluValGlyGlu 255
DB 1245 GGCAATCCGCTATGAAACCAAAATGTCCTTGAGTAGACATATCAAAAAGATTGGAGAA 1304
QY 256 AAlserIleIysValProGlnLeuIysMeGluIle 267
DB 1305 GCTTCCATTAAGTACCAATTAATAATGGAGATA 1340

RESULT 13
US-10-012-140-6
; Sequence 6, Application US/10012140
; Publication No. US2003009017A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USBS
; FILE REFERENCE: 381552004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-012-140-6

Alignment Scores:
Pred. No.: 1,05e-164 Length: 1245
Score: 1348.00 Matches: 260
Percent Similarity: 98.50% Conservative: 3
Best Local Similarity: 97.38% Mismatches: 4
Query Match: 97.33% Indels: 0
DB: Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-012-140-6 (1-1245)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIysValGluPheProPheLeuVal 20
DB 442 ATGAGGCTCATGACCTTACTATTAGGTGACCAATTAAGTGAATTTCTTTTAAAGTT 501
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 502 CTTTGTATTTCTGGAGGTCACGTGCTGTGTGCAATTAGTTCAAGAGATTTCATTTTCG 561
```



Qy	41	LeuLeuEnGIyLysSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArg	60
Db	562	CTTCCTGGAAAGCTCTTGGACATAGCACAGGTGACATCTTGCACAGGTGGCAGAA	621
Qy	61	LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaLeuHis	80
Db	622	CTTTCCTTAAATAAACATCCAGAGTGCTCCACATAGTGAGTGGAAACCTATGACAT	681
Qy	81	LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys	100
Db	682	TTGGCCAAACAGGAATATGATTTGATTTGACATCAACCTCCCTTGATCATGCTAA	741
Qy	101	AsnGlyAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleLeuLysLys	120
Db	742	AAATGTGATTTTCTTTTAACTGGACTCAACGTTTACGTAAATAATATGAAAAACAG	801
Qy	121	GluLysGlnGluGlnGlyIleGluLysGlnGlnIleLeuSerSerAlaAlaAspIleAlaI	140
Db	802	AAACAGAGGAGATGATTTGAGAAAGGGGCAAAATCTGTCTTCAGCAGACATTCCTGC	861
Qy	141	ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe	160
Db	862	ACAGTAGCACACACATGSCATGTATCTGTGAAAGAACATCGGGCTATTCGTTT	921
Qy	161	CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal	180
Db	922	TGTAAAGCAGAGAGACTTGTTCCTCAAAATATATGCGATCTGGTTCATCTGGTGTGC	981
Qy	181	AlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr	200
Db	982	GCAAGTAACTTCTAATCCGACAGAGCTCTGGAAATTTTAAACAAACGCAACACGTGCACCT	1041
Qy	201	LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTPraAsnGly	220
Db	1042	TTGTTGTGTCCTCTCCACGACTTATGCACTGATATAGCATTTATGATTCATGAGATGGT	1101
Qy	221	IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu	240
Db	1102	ATTGAAAGACTACGTGCTGGCTTGGCACTTTTATCATGACATGAAAGGCAATCCGCTATGAA	1161
Qy	241	ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal	260
Db	1162	CCAAATATGCTCTCTTGAGATAGACATATCAAAAAGAGTTGGAGAACCTTCCATAAAAAGTA	1221
Qy	261	ProGlnLeuLysMetGlnIle 267	
Db	1222	CCACAATTAAAAATGAGATA 1242	

```

RESULT 14
US-10-012-140-4
; Sequence 4, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 381552004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1820
; TYPE: DNA

```

```

? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (146)...(1390)
? PEPTIDE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(1820)
? OTHER INFORMATION: n = A,T,C or G
US-10-012-140-4

```

**Alignment Scores:**

Pred. No.:	1,856-164	Length:	1820
Score:	1348.00	Matches:	260
Percent Similarity:	98.50%	Conservative:	3
Best Local Similarity:	97.38%	Mismatches:	4
Overl. March:	97.33%	Indels:	0

DB:

Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) X US-10-012-140-4 (1-1820)

Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIysValGluPheProPheLeuVal 20

Db 587 ATGAGGCTCATGCACTTACTATTAGGTGACCAATAAGTAGAATTCCTTTTTAGTT 64

21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40

647 CTTTTCATTTCAGAGTCACTGTCTGTTGCACTTACGTTCAAGGAGTTTCAGATTTCG 70

1. The first step is to identify the problem or question that needs to be answered. This involves understanding the context and the specific requirements of the task.

Age	Percentage
18	10
19	10
20	10
21	10
22	10
23	10
24	10
25	10
26	10
27	10
28	10
29	10
30	10
31	10
32	10
33	10
34	10
35	10
36	10
37	10
38	10
39	10
40	10
41	10
42	10
43	10
44	10
45	10
46	10
47	10
48	10
49	10
50	10
51	10
52	10
53	10
54	10
55	10
56	10
57	10
58	10
59	10
60	10
61	10
62	10
63	10
64	10
65	10
66	10
67	10
68	10
69	10
70	10
71	10
72	10
73	10
74	10
75	10

DB /0/ C1C1GGAGGAC11GGACCA1AGCACAG1GAACA1GC1GGACAAAG1GGCCAAAGAAAG /0

61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGluHis 80

Db 767 CTTCTTAATAACATCCAGAGTGCTCCACCATGAGTGTGGAAAGCCATAGAACAT 82

81 LeuAlaLysGlnGlyAsnArgPheHisIvPheAspIleLysProPheHisAlaLys 10

Db 827 TTGGCCAAACAGGAATAGATTTCATTTTGACATCAACCTCCCTTGATCATGCTAA 88

101 AcnNcAcnDhScerPheThrc[vlauG]nh[ava]ThrAsnI[st]et]eMetI[st]Iys 12

[illegible][illegible][illegible]

Db 947 AAACAAGGAGGTATTGAGAAGGGCAATCCTGTCTTCAGCAGCAGACATTGCTGCC 100

141 ThrValGlnHisThrMetAlaCysHisIleuValLysArgThrHisArgAlaIleLeuPhe 16

Db 1007 ACAGTACAGCACACAATGGCATGTCATCTTGTGAAGAACAACATCGGCTATTCTGTTT 10

161 Cvatwsg]narcAsnl,euleuProglnasnAsna]ava]leuyajAlaserc]vg]vva] 18

[illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

101 A1A0C1B1B1F1E1Y 111E1N1G1H1I1J1K1L1M1N1O1P1Q1R1S1T1U1V1W1X1Y1Z1

D5 1127 GCAAGTAACTTCTATATCCGACAGAGCCTCTGGAAATTTTAACTAAACGCAACACAGTGCAT 11

QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 22

Db 1187 TTGTTGTCTCTCCAGACTATGCACCTGATAATGGCATTATGATTCATGGAATGGT 12

221 IlegIuArQLeuArqAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 24

1347 ATTCGACCTACCTGCTTGGGCTTTTACATGACATAGAGGCATCCGCTATGAA 13

**Figure 1**

[illegible]

D5 130 / CCAAAAGTCTCTTGGAGTAGACAATCAAAAGGATGTTGGAGGAGCTTCCATATTTGATG

261 Proglutiny met Glutile 267

Db 1367 CCACAAATTAATAATGAGATA 1387

RESULT 15  
US-10-094-749-400  
Sequence 400. Application US/10094749  
Publication No. US20030219741A1

GENERAL INFORMATION:  
APPLICANT: ISOGAI, TAKAO  
APPLICANT: SUGIYAMA, TOMOYASU  
APPLICANT: OTSUKI, TETSUJI  
APPLICANT: WAKAMATSU, AI  
APPLICANT: SATO, HIROYUKI  
APPLICANT: ISHII, SHIZUKO  
APPLICANT: YAMAMOTO, JUN-ICHI  
APPLICANT: ISONO, YUUKO  
APPLICANT: HIO, YURI  
APPLICANT: OTSUKA, KAORU  
APPLICANT: NAGAI, KEIICHI  
APPLICANT: IRIE, RYOTARO  
APPLICANT: TAMECHIKA, ICHIRO  
APPLICANT: SEKI, NAOHICO  
APPLICANT: YOSHIKAWA, TSUTOMU  
APPLICANT: OTSUKA, MOTOTYUKI  
APPLICANT: NAGAHARI, KENJI  
APPLICANT: MASUHO, YASUHIKO  
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA  
FILE REFERENCE: 084335/0160  
CURRENT APPLICATION NUMBER: US/10/094,749  
CURRENT FILING DATE: 2002-03-12  
PRIOR APPLICATION NUMBER: 60/350,435  
PRIOR FILING DATE: 2002-01-24  
PRIOR APPLICATION NUMBER: JP 2001-328381  
PRIOR FILING DATE: 2001-09-14  
NUMBER OF SEQ ID NOS: 3381  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 400  
LENGTH: 2208  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-094-749-400

Alignment Scores:  
Pred. No.: 1,47e-145 Length: 2208  
Score: 1204.00 Matches: 239  
Percent Similarity: 89.51% Conservative: 0  
Best Local Similarity: 89.51% Mismatches: 4  
Query Match: 86.93% Indels: 24  
DB: 18 Gaps: 1

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-094-749-400 (1-2208)

QY 1 MetGluA1aH1sAlaLeuThr11eArGLeuThraenLySValGluPheProPheLeuVal 20  
Db 785 ATGGAGGCTCATGCACTTAATTAGGTGACCAATAAGTAGAATTCCTTTTAACT 844

QY 21 LeuLeu11eSerG1yG1yH1sCyLeuLeuA1aLeuValG1nG1yValSerAspPheLeu 40  
Db 845 CTTTGAATTTCTGGAGGCTGCTGCTGTGGCATTAGTTCAAGAGTTTCAGATTTTCG 904

QY 41 LeuLeuG1yLySerLeuAsp11eAlaProG1yAspMetLeuAspLySValAlaArgArg 60  
Db 905 CTTCTTGAAGAAGTTTGGACATAGCACCAAGTGAATGCTTGACAGGTGGCAAGAA 964

QY 61 LeuSerLeu11eLySH1sProG1uCySerThrMetSerG1yG1yLySAla11eG1uH1s 80  
Db 965 CTTCTTTAATAAACAATCCAGAGTCTCCACATGATGGTGGGAAAGCCATAGACAT 1024

QY 81 LeuA1aLySG1nG1yAsnArgPheH1sPheAsp11eLySProProLeuH1sH1sAlaLyS 100  
Db 1025 TTGGCCAAACAAAGAAATAGATTTCATTTTGAATCAATCAAAACCTCCCTGCATCATCTAAA 1084

QY 101 AsnCysAspPheSerPheThrG1yLeuG1nH1sValThraSplyS11e1eMetLyS1yS 120

Db 1085 AATGTGAATTTCTCTTTACTGACTTCAACGCTTAATGATAAATTAATGAAG 1144

QY 121 G1uLySG1uG1y11eG1uLySG1yG1n11eLeuSerSerA1a1aAsp11eA1a1a 140

Db 1145 GAAAAAGAGAGATTTTGAAGAGGGCAATCTCTCTTCAGACACACATTCGTGCC 1204

QY 141 ThrValG1nH1sThrMetA1aCySH1sLeuValLySArgThrH1sArgA1a11eLeuPhe 160

Db 1205 ACAGTACACACACAAATGCGCATGTCATCTTGGAAGAAACACATCGGGCTATTCTGTTT 1264

QY 161 CysLySG1aArgAspLeuLeuProG1naAsnA1aValLeuValA1aSerG1yG1yVal 180

Db 1265 TGTAAAGCAGAGAACTGTGTACCTCAAAATATGCACTGTGTCATCTGGTGGTGC 1324

QY 181 AlaSerAspPheTyrl1eArgArgA1aLeuG1u11eLeuThraSna1aThrG1nCySThr 200

Db 1325 GCAAGTAACTTCTGTAATCCGCAAGCTCTGGAAATTTTAACAAACGCAACAGTGCAT 1384

QY 201 LeuLeuCySProProProArgLeuCySThrAspAsnG1y11eMet11eA1aThraSng1y 220

Db 1385 TTGTTGTGTCCTCTCCAGACTATGCACTGATTAATGGCATTATGATTGCA----- 1435

QY 221 11eG1uArgLeuArgA1aG1yLeuG1y11eLeuH1sAsp11eG1uG1y11eArgTyG1u 240

Db 1435 ----- 1435

QY 241 ProLySAspProLeuG1yValAsp11eSerLySG1uValG1yG1uA1aSer11eLySVal 260

Db 1436 ----TGATGTCTCTTGGAGTAAACATATCAAAAGATTGAGAGACTTCATTAAGTA 1492

QY 261 ProG1nLeuLySmetG1u11e 267

Db 1493 CCACAAATTAATAATGAGATA 1513

Search completed: November 11, 2005, 03:16:33  
Job time : 896.614 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: November 10, 2005, 16:54:39 ; Search time 189.512 Seconds  
(without alignments)  
2305.321 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414

Perfect score: 1385  
Sequence: 1 MEHALTIRLTKVPEPPLV.....DISKVGASIKVQLKMEI 267

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=x1p  
-Q=/cgn2\_1/USFTO.spool\_p/US10649273/runat\_02112005\_091339\_15581/app.query.fasta\_1.1429  
-DB=Issued\_Patents\_NA -OPT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=100 -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10649273 @CGN 1.1 177 @unat 02112005 091339 15581 -NCPU=6 -ICPU=3  
-NO MAP -LARGEOUTRY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MAIN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : Issued Patents NA:\*

1: /cgn2\_6/ptcdat/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptcdat/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptcdat/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptcdat/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptcdat/1/ina/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptcdat/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1385	100.0	1416	4	US-09-774-528-177
2	1385	100.0	1526	4	US-10-067-443-23
3	1385	100.0	2197	4	US-10-067-443-1
4	1385.5	98.1	1387	4	US-10-067-443-21
5	824.5	59.5	14364	4	US-10-067-443-20
6	308.5	22.3	94750	4	US-09-596-002-38
7	307	22.2	1053	4	US-09-540-236-806
8	291.5	21.0	1059	4	US-09-252-991A-884
9	291.5	21.0	1256	4	US-09-252-991A-801
10	289.5	20.9	1830121	4	US-09-557-884-1
11	289.5	20.9	1830121	4	US-09-643-990A-1
12	282	20.4	996	4	US-09-902-540-6612

C	13	282	20.4	2582	4	US-09-902-540-503	Sequence 503, App
	14	280.5	20.3	1074	4	US-09-543-681A-2341	Sequence 2341, App
	15	270.5	19.5	1315	1	US-08-087-797-1	Sequence 1, Appli
	16	268.5	19.4	1032	4	US-09-489-039A-2050	Sequence 2050, App
	17	261	18.8	1092	4	US-09-107-532A-2955	Sequence 2955, App
	18	254	18.3	1008	3	US-08-987-121A-5	Sequence 5, Appli
	19	254	18.3	1011	3	US-08-987-121A-3	Sequence 3, Appli
	20	250	18.1	1006	4	US-09-536-784-51	Sequence 51, Appli
C	21	250	18.1	10974	3	US-08-961-527-214	Sequence 214, App
	22	248	17.9	1011	3	US-09-066-512-1	Sequence 1, Appli
	23	246	17.8	1011	4	US-09-583-110-2196	Sequence 2196, App
	24	241	17.4	1011	4	US-09-107-433-1618	Sequence 1618, App
	25	236.5	17.1	640681	4	US-09-790-988-1	Sequence 727, Appli
	26	236	17.0	822	4	US-09-710-279-727	Sequence 727, App
	27	236	17.0	1101	4	US-09-134-000C-1551	Sequence 1551, App
C	28	236	17.0	3993	4	US-09-134-000C-1551	Sequence 3985, App
	29	236	17.0	3993	4	US-09-134-000C-1551	Sequence 1072, App
	30	233	16.8	1107	3	US-09-149-624-1	Sequence 1, Appli
	31	230	15.9	1026	3	US-09-149-624-1	Sequence 794, App
C	32	218	15.7	3064	3	US-09-221-017B-794	Sequence 1, Appli
	33	218	15.7	123025	4	US-09-198-452A-1	Sequence 102, App
	34	218	15.7	1230230	4	US-09-438-185A-1	Sequence 102, App
C	35	216	15.6	15249	4	US-08-956-171E-102	Sequence 102, App
	36	216	15.6	15249	4	US-08-781-986A-102	Sequence 102, App
C	37	215.5	15.6	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	38	215.5	15.6	4411529	3	US-09-103-840A-1	Sequence 147, App
C	39	211.5	15.3	1155	4	US-09-602-777A-147	Sequence 1, Appli
	40	207.5	15.0	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	41	207.5	15.0	1664976	4	US-09-692-570-1	Sequence 6, Appli
	42	192.5	13.9	1663	4	US-09-620-312D-6	Sequence 131, App
C	43	185.5	13.4	42325	4	US-08-311-731A-121	Sequence 261, App
	44	181.5	13.1	432	4	US-09-328-352-261	Sequence 19, Appli
	45	169	12.2	876	4	US-09-724-623-19	

## ALIGNMENTS

RESULT 1  
US-09-774-528-177  
Sequence 177, Application US/09774528 P.D.  
Patent No. 6743619  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Zhou, Ying  
APPLICANT: Goodrich, Ryle  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Ren, Feiyang  
APPLICANT: Zhang, Jie  
APPLICANT: Zhao, Qing A.  
APPLICANT: Yang, Yonghong  
APPLICANT: Xue, Aidong J.  
APPLICANT: Wehrman, Tom  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Wang, Dunrui  
APPLICANT: Dymnac, Radoje T.  
TITLE OF INVENTION: No. 6743619el Nucleic Acids and  
TITLE OF INVENTION: Polypeptides  
FILE REFERENCE: 802  
CURRENT APPLICATION NUMBER: US/09/774,528  
CURRENT FILING DATE: 2001-01-30  
NUMBER OF SEQ ID NOS: 441  
SOFTWARE: pt\_fl\_genes Version 2.0  
SEQ ID NO 177  
LENGTH: 1416  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (205) ..(1305)  
US-09-774-528-177  
Alignment Scores:

Pred. No.: 2,516-177 Length: 1416  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-09-774-528-177 (1-1416)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeuValGluPheProPheLeuVal 20  
DB 502 ATGAGGCTCATGACCTTACTATTAGGTGACCAATTAAGTAGAATTCCTTTTATGTT 561  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 562 CTTTGAATTTCTGAGGCTCATGCTGTGGCATTAGTTCAAGAGGTTTCAGATTTTCTG 621  
QY 41 LeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
DB 622 CTTCTGAAAGTCTTTGACATAGCACAGGTGACATGCTTGCACAGGTGGCAAGAAGA 681  
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyIleValAlaIleGlnHis 80  
DB 682 CTTTCTTAATAAACAATCCAGAGTCTCCACCATGAGTGTGGAAAGCCATTAGACAT 741  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
DB 742 TTGGCCAAACAGAAATAGATTTCATTTTGCATCAAACTCCCTTGATCATGTCTAA 801  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValIleAspLysIleIleMetLysLys 120  
DB 802 AATTGTGATTTTCTTTTACTGACCTTCAACGTTTACGATTAATATATATGAAGAAG 861  
QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140  
DB 862 GAAAGAGGAAAGATAGAGAGGGGCAAAATCCTGTCTTCAGACAGACATTCCTGTC 921  
QY 141 ThrValGlnHisThrMetAlaCysHisIleuValLysArgThrHisArgAlaIleLeuPhe 160  
DB 922 ACAGTACAGCACACATGAGCATGTCATCTTGAAAGAAACACATCGGGCTATTCGTT 981  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyIleVal 180  
DB 982 TGTAGACAGAGACTTCTTACTCTCAAAATATAGACATCTGTTGCATCTGTGTGTC 1041  
QY 181 AlaSerAsnPheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
DB 1042 GCAAGTAACTTCTATATCCGACAGCTCTGGAAATTTTAAACAACGCAACAGTGCAT 1101  
QY 201 LeuLeuCysProProArgLeuLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 220  
DB 1102 TTGTGTGTCTCTCCCTCCAGACTATGCACTGATTAATGGATTATATGCAATGATGT 1161  
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrglu 240  
DB 1162 ATTGAAGACTACGCTGCTGGCATTTTACATGACATGAAGAGCATCCGCTATGA 1221  
QY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260  
DB 1222 CCAAAATGCTCTCTGGAGTAGACATATCAAAAGAAAGTTGAGAGCTTCATTAAGA 1281  
QY 261 ProGluLeuLysMetGluIle 267  
DB 1282 CCACAAATTAATAATGAGATA 1302

RESULT 2  
US-10-067-443-23

; Sequence 23, Application US/10067443  
; Patent No. 6642041  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; TITLE OF INVENTION: SPINAL CORD, ME-1

; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: Patent In version 3.0  
; SEQ ID NO 23  
; LENGTH: 1526  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-067-443-23

Alignment Scores:  
Pred. No.: 2,856-177 Length: 1526  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-067-443-23 (1-1526)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeuValGluPheProPheLeuVal 20  
DB 1 ATGAGGCTCATGACCTTACTATTAGGTGACCAATTAAGTAGAATTCCTTTTATGTT 60  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 61 CTTTGAATTTCTGAGGCTCATGCTGTGGCATTAGTTCAAGAGGTTTCAGATTTTCTG 120  
QY 41 LeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
DB 121 CTTCTGAAAGTCTTTGACATAGCACAGGTGACATGCTTGCACAGGTGGCAAGAAGA 180  
QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140  
DB 141 ThrValGlnHisThrMetAlaCysHisIleuValLysArgThrHisArgAlaIleLeuPhe 160  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyIleVal 180  
DB 181 AlaSerAsnPheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
DB 301 AATTGTGATTTTCTTTTACTGACCTTCAACGTTTACGATTAATTAATGAAGAAG 360  
QY 301 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 400  
DB 361 GAAAGAGGAAAGATAGTGAAGAGGGGCAAAATCCTCTTCACAGACAGCATTCGCTGC 420  
QY 421 ThrValGlnHisThrMetAlaCysHisIleuValLysArgThrHisArgAlaIleLeuPhe 160  
DB 481 TGTAGACAGAGACTTGTACTCTCAAAATTAAGCAGTACGTGTGCATCTGTGTGTC 540  
QY 541 GCAAGTAACTTCTATATCCGACAGCTCTGGAAATTTTAAACAACGCAACAGTGCAT 600  
QY 601 LeuLeuCysProProArgLeuLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 220  
DB 601 TTGTGTGTCTCTCCCTCCAGACTATGCACTGATTAATGGATTATGATGAATGATGT 660  
QY 661 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrglu 240  
DB 661 ATTGAAGACTACGCTGCTGGCATTTTACATGACATGAAGAGCATCCGCTATGA 720



```
Db 645 CTTTCTTTAAATAAACAATCCAGAGCTCCACCAAGATGGTGGGAAAGCCATTAGACAT 704
Qy 81 leuAlaIysGInGIyAsnArgPheHisPheAspIleuYsProLeuHisIleAlaIys 100
Db 705 TTGGCCAAACAGAAATAGATTTCATTTTGACATCAAACTCCCTTCGATCATCTTAA 764
Qy 101 AsnCyAspPheSerPheThrGIyLeuGInHisValIThrAspLysIleIleMetLysIys 120
Db 765 AATTGTGATTTTCTTTTACTGACTTCAACACGTTACTGATTAATAATATGAAAAAG 824
Qy 121 GIuIysGIuGIuGIyIle----- 126
Db 825 GAAAGAGGAGGATATTTCTATATTAGTAAAGTTGAACAGATTAATTCCTGGAATG 884
Qy 127 -----GIuIysGIuGIuIleIleuSerSerIa 135
Db 885 TGCCATAAAATAGCTGCTCATTTTTCGACAGTATGAAAGGGCAAAATCTGCTTCAGCA 944
Qy 136 AlaAspIleAlaIaIthrValGIuInHisPheMetAlaCySHIsIleuValIysArgThrHis 155
Db 945 GCACACATTCGTCGCACAGTACAGACACAAATGSCATGTCATCTTGTAAGAAAGACAT 1004
Qy 156 ArgAlaIleIeuPheCySlySGInArgAspIleuLeuProGIuAsnAsnAlaValIleuVal 175
Db 1005 CGGGCTATTCGTTTGTGAACAGAGACATGTTTACCTCAAAATTAAGCAGTACTGTT 1064
Qy 176 AlaSerGIyGIuValAlaSerAsnPheTyrlIeaArgArgAlaIleuGIuIleuThrIen 195
Db 1065 GCATCTGTGTGTGTGCAAGTAACTTCTATATCCCAAGCTCTGAAATTTTAAACAAC 1124
Qy 196 AlaThrGIuCySThrLeuLeuCySProProProArgLeuCySThrAspAsnGIyIleMet 215
Db 1125 GCACACAGTGCATCTTTTGTGTCTCTCCACACATGACATGATATGCGCATTAAG 1184
Qy 216 IleAlaTrpAsnGIyIleGIuArgLeuArgAlaGIyIleuGIyIleuHisAspIleGIu 235
Db 1185 ATTCATGAGATGGATGAAAGAACTACGTGTGCTGGGCACTTTTACATGACATAGAA 1244
Qy 236 GIyIleArgTyrgIuProIyCySProLeuGIyValAspIleSerIyGIuValGIyGIu 255
Db 1245 GGCATCCCTATGAAACCAAAATGTCCTCTTGAGATGACATATCAAAAGAGTTGAGAA 1304
Qy 256 AlaSerIleIysValProGIuLeuIysMetGIuIle 267
Db 1305 GCTTCATTAAGTACCAATTTAAATAATGAGATA 1340

RESULT 5
US-10-067-443-20
; Sequence 20, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: SPINAL CORD, MP-1
; CURRENT APPLICATION NUMBER: D0073 NP
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 14364
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-20

Alignment Scores:
Pred. No.: 1,346-99 Length: 14364
Score: 824.50 Matches: 240
```

```
Percent Similarity: 26.73%
Best Local Similarity: 26.73%
Query Match: 59.53%
DB: 4

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-20 (1-14364)

Qy 1 MetGIuAlaHisAlaIleuThrIleArgLeuThrAsnLysValGIuIleProPheLeuVal 20
Db 10845 ATGAGGCTCATGCACTTACTATTAAGTTGACCAATTAAGTAAGTAATTTCTTTTAAATT 10904
Qy 21 LeuIleuIleSerGIyGIyHisCySLeuLeuAlaIleuValGIuGIyValIserAspPheLeu 40
Db 10905 CTTTGAATTTTCGAGGTCACGTCGTGTGGCAATTAGTTCAGAGAGTTTCAGATTTTCTG 10964
Qy 41 LeuIleuGIyIysSerLeuAspIleAlaProGIyAspMetLeuAspLysVal 57
Db 10965 CTTTCATGAAAGTCTTTGACATAGCACCAAGTGACATGCTTGACAAAGGT-AAATTAAAGAA 11023
Qy 57 ----- 57
Db 11024 TTAATTTCTCATCTTTTGTGTAATGTTCATTTCACTAAGTAGACATATGATGTGC 11083
Qy 57 ----- 57
Db 11084 TACCAACATTCACCTAAATATTTCTGAATTTTATCTTAGTAAACGAAAAAATTCACAT 11143
Qy 57 ----- 57
Db 11144 ATGTGAGAAAAAATAGAAAGTAGTACACAAATTTATATTCTTAGCCTTTCTTAATA 11203
Qy 57 ----- 57
Db 11204 AAATGTAAAGAGTTCATATCTGTACATTAAGAGCTGAATATGTTGACATATACATTATG 11263
Qy 57 ----- 57
Db 11264 TATTTTGCCAAATAAATGTATGTGAAAAAGAGTGCCTGTAAACTTAACATACGAAAAAA 11323
Qy 57 ----- 57
Db 11324 GGTAAATTAAGAAATATATATATATATTAATTAACATTAAGACATTAAGATGCAATGACAGAA 11383
Qy 57 ----- 57
Db 11384 TTAATACACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11443
Qy 57 ----- 57
Db 11444 AATTAATACAGAGGCTACTGCAATATATAGAAAAAACAACAAACAACAAACAACACTG 11503
Qy 57 ----- 57
Db 11504 CTTCCACAGTAAATTAATTAAGAAATTAAGAAAGTCTTATTAATTAAGCTTCATCAATT 11563
Qy 57 ----- 57
Db 11564 AAGCAGTATTTGTCAACTTCAAGCCATTTTCCAAACATTAAGAAAGAAACAAATAGACAG 11623
Qy 57 ----- 57
Db 11624 GGGCAGTATGGGCTCTTATTTGTCGGGTCAATATAGGAACAGGTTGTCTGTACG 11683
Qy 57 ----- 57
Db 11684 TGAATATCAGCTATAGCTATATTTTGCACAAAGTATAGACATGTTTATTCATTCAGGGGTT 11743
Qy 57 ----- 57
Db 11744 TTTTGTGTTGTTAGTAATTTTCAATTTATTTCTTTGACATCTTTGTTTCACAGTATTT 11803
Qy 58 -----AlaArgArgLeuSerLeuIle 64
```

Db 11804 AATTATGACTCTAAAAAATATGTTCTTTCATAGTCGCAAGAACTTTCTTAATA 11863  
QY 65 LysHisProGluCySerThrMetSerGlyIyLysAlaIleGluHisLeuAlaIyGln 84  
Db 11864 AAACATCCAGAGTCTCCACCATGAGTGAGGGAAGCAATGAAACATTGGCCAAACAA 11923  
QY 85 GlySerMetArgPheHisPheAspIleIySerProProluHisIleAlaIySerGlySerPhe 104  
Db 11924 GGAATATGATTTCAATTTTGACATCAACCTCCCTGCATCATGCTAAATAATGTGATTTT 11983  
QY 105 SerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLeuGlyLysGluGln 124  
Db 11984 TCTTTTACTGACATTCACACGTTACTGATATAATATATGAAAAAGAAAAAGAGGAA 12043  
QY 124 ----- 124  
Db 12044 GGTATATTTCTAATATGTAAGATGACAGATTAATATCTCGATGTCCTAAAAATA 12103  
QY 125 -----GlyIleGlyLysGlyGlnIleLeuSerSerAlaIleAspIleAl 139  
Db 12104 GCTGCTCATTTCTGCAAGGTATTTGAGAAAGGGCAATCTGCTTCAGCGCAGCAATGTC 12163  
QY 139 AlaIleThrValGlnHisThrMetAlaCyHisIleuValIySerGlyThrHisArgAlaIleLe 159  
Db 12164 TGCACAGTACAGACACAAATGCGATGTCATCTTGTAAGAAAGACACATCGGGCTATTTCT 12223  
QY 159 uPheCySerGlyLeuIyArgSerLeuLeuProGlnAsnAsnAlaValLeu----- 174  
Db 12224 GTTTTGTAGAGAGAGACTGTGTACTCTCAAAATATATGACAGTACTGTATAGTTTATCTC 12283  
QY 174 ----- 174  
Db 12284 ATTATATAGTATATGTTACACTTTGCAATATGTTACTTTTCCCAAGACTTGACCTTG 12343  
QY 174 ----- 174  
Db 12344 TGTTTAGATGACAGATCTTTATGCTTATGCTTATGCCCCGACAGTATGAAATATATGACAG 12403  
QY 174 ----- 174  
Db 12404 GATAGAAAAGACTAACGCCATTTCTGTACTAGTTTGATGCTTTATGGAACAGCTGTA 12463  
QY 174 ----- 174  
Db 12464 TAGCTTCTATGACATTAAGTCTAATTTTGCACTCTTCTGTGATTTAAAGAGGGCTT 12523  
QY 174 ----- 174  
Db 12524 ACAATAAAGAAAGTAAATGACATGCTATGCTATCTATTTTAAAGAAATATAGTGCATTT 12583  
QY 174 ----- 174  
Db 12584 CTTTCATCTTTGATGAATCCCTTTGTTGTTGTTTAAATGACAGCTCAAAATTT 12643  
QY 174 ----- 174  
Db 12644 AGCAGTGGAGGTGATTTCCAACTTTGCTGACACTAATGTTGATTAAGTTCTGATAATC 12703  
QY 174 ----- 174  
Db 12704 CACTATATTTGACAGCCAAATCCCTTAATATGCTTAAAGGCTTGACAAACATCT 12763  
QY 174 ----- 174  
Db 12764 GTTTACTGTATCTTAACTTTATCTTAAATAATTAATACTAAAGTGGGAAATGTT 12823  
QY 174 ----- 174  
Db 12824 TAAATGAGTAAATTCATAGATGGAATTTTACATGGAATATCAAGAAATATTTTTCAGA 12883  
QY 174 ----- 174  
Db 12884 GTTATGATGTAATAATGCAAAATATTAATAAATTTGAGGGTCTAATAATATGCTACTATGA 12943

QY 174 ----- 174  
Db 12944 TTGAATATATTAATAATAATATTTAGATGAAAGTTGGAAGAAATATACAAAATGCT 13003  
QY 174 ----- 174  
Db 13004 AGTAAATGTTGATGCTATTAAGATTAATTAATTTTCTTCCAAATTTTATATAC 13063  
QY 174 ----- 174  
Db 13064 ATAGATATGTCATCTGCCCATTAACCATCTCAAAATGGGATAGTTATTTATGTTAATG 13123  
QY 174 ----- 174  
Db 13124 CTGATATTTTCTCCAGTTTAATAGCAGCTTGTTCAATCCATATATGATGATTAAT 13183  
QY 175 -----ValAlaSerGlyGlyValAlaSerAspPheTyrIleAr 187  
Db 13184 TTGGTTTCTCAATTCCTTCAGGTGTCATCTGTGTGTGTCGAAGTAACTTCTATATCCG 13243  
QY 187 GATGAlaLeuGluIleLeuThrAsnAlaThrGlnCySerThrLeuCySerProProAr 207  
Db 13244 CAGAGCTCTGAAATTTTAAACAAACGCAACACAGTCACTTGTGTGCTCTCCAG 13303  
QY 207 gluCySerThrAspAsnGlyIleMetIleAlaTTP----- 218  
Db 13304 ACTATGCACTGATTAATGCAATTAATGATTCATGTAAGCCACAGATATACGCTTAC 13363  
QY 218 ----- 218  
Db 13364 TCATACTATGTAATAATTAATTTGCACTTTTATCATACTAAGCCTTCTTCAGATCT 13423  
QY 219 -----AsnGlyIleGluArgLeuArgA 226  
Db 13424 TGGAGCTATGATTTTATTAATTAATGCTTCTTAATTAAGAAATGATTAAGAAAGCTACGTG 13483  
QY 226 IAGIyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProLys 242  
Db 13484 CTGGCTTGGGCACTTTTACATGACATAGAAAGCATCCGCTATGAAACCAAG 13533

RESULT 6  
US-09-596-002-38  
; Sequence 38, Application US/09596002  
; Patent No. 6632636  
; GENERAL INFORMATION:  
; APPLICANT: Lagace, Robert, E.  
; APPLICANT: Paterson, Chandra  
; APPLICANT: Berg, Kam, L.  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME  
; FILE REFERENCE: PM-0008-4 US  
; CURRENT APPLICATION NUMBER: US/09/596, 002  
; CURRENT FILING DATE: 2000-06-16  
; PRIOR APPLICATION NUMBER: 60/140,121  
; PRIOR FILING DATE: 1999-06-18  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PERL Program  
; SEQ ID NO 38  
; LENGTH: 94750  
; TYPE: DNA  
; ORGANISM: M. catarrhalis  
; FEATURES:  
; NAME/KEY: misc. feature  
; OTHER INFORMATION: incyte template ID No. 6632636 38  
; PUBLICATION INFORMATION:  
US-09-596-002-38

Alignment Scores:  
Pred. No.: 3,21e-28 Length: 94750  
Score: 308.50 Matches: 87  
Percent Similarity: 49.61% Conservative: 40  
Best Local Similarity: 33.98% Mismatches: 101  
Query Match: 22.27% Indels: 29





```
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 884
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-884

Alignment Scores:
Pred. No.: 3,67e-29      Length: 1059
Score: 291.50           Matches: 87
Percent Similarity: 50.41% Conservative: 37
Best Local Similarity: 35.37% Mismatches: 99
Query Match: 21.05%     Indels: 23
                        Gaps: 9
DB:

US-10-649-273-2_COPY_148_414 (1-267) x US-09-252-991A-884 (1-1059)
QY      1 MetGluAlaHisAlaLeuThrIleArgLeuThraAsnLys---ValGluPheProPheLeu 19
Db      711 ATGGAAGGGCACCTGCTGCGCCGCGATGCTGGAAAGACGCCACCGCGGTTCCCGTTCCGTC 652
QY      20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPhe 39
Db      651 GCCTTGCGTGGTTCCGGCGGTCAACCCAGTTGGCGGGTGGAGGATGCGCGCCGCTAC 592
QY      40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspIlyValAlaArg 59
Db      591 CAGTTGCTGGCGAATCGGTGACGATGCCCGCGCGAAGCCTTGACAAAGCCGCAAG 532
QY      60 ArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGlu 79
Db      531 CTGATCGGCGCTG---GGCTATCCC-----GGTGTGTCGGAAATCGCC 493
QY      80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAla 99
Db      492 CGCTGGCGGAGCGCGGACACTCCCTGCGCTTCGTTCCCGCGGCCGATGACCGATGCC 433
QY      100 LysAsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119
Db      432 CCGCGCTGAGACTTCAGCTTCAGCGGGCTCAAGACCTTACCTTAAC---ACCTGGCAG 376
QY      120 LysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 139
Db      375 CGTTGCGTGGAGCGCGGACGACGACGAGCAG-----ACCGCTGGCAGATCGCC 325
QY      140 AlaThrValGlnHisThrMetAlaCysHisIleValLysArgThrHisArgAlaIleLeu 159
Db      324 CTGGCGTTCCAGACCGCGGCTGTGAGACCTGCTGATCAAGTGCCTGCGCCCTTG--- 268
QY      160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyGly 179
Db      267 -----AAGCAGACCGGCTG-----AAGAAC-----CTGTATATCGCGCGGT 229
QY      180 ValAlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCys 199
Db      228 GTACGCCCAACAGAGCGCTGCGCAGCGGCTGGAAGAGATGCTCGCGAATGAAGAGGG 169
QY      200 ThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsn 219
Db      168 CAGGTGTCTTACGCCCGCGCGCTTCTGCACCGACATGCGCGCATATATGCCCTAACGCC 109
QY      220 GlyIleGluArgLeuArgAlaGlyIleuGlyIleuHisAspIleGluGlyIleArgTyx 239
Db      108 GGTGCGCAGCGCTGTCTGCGCGG-----CAGCATGAAGCGCGCGCGATGACGGCTC 58
QY      240 GluProLysCysProLeu 245
Db      57 CAGCGCGCGCTGCGCGATG 40
```

```
RESULT 9
US-09-252-991A-801
; Sequence 801, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 801
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-801

Alignment Scores:
Pred. No.: 4.55e-29      Length: 1206
Score: 291.50           Matches: 87
Percent Similarity: 50.41% Conservative: 37
Best Local Similarity: 35.37% Mismatches: 99
Query Match: 21.05%     Indels: 23
                        Gaps: 9
DB:

US-10-649-273-2_COPY_148_414 (1-267) x US-09-252-991A-801 (1-1206)
QY      1 MetGluAlaHisAlaLeuThrIleArgLeuThraAsnLys---ValGluPheProPheLeu 19
Db      514 ATGGAAGGGCACCTGCTGCGCCGCGATGCTGGAAAGACGCCACCGCGGTTCCCGTTCCGTC 573
QY      20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPhe 39
Db      574 GCCTTGCGTGGTTCCGGCGGTCAACCCAGTTGGCGGGTGGAGGATGCGCGCCGCTAC 633
QY      40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspIlyValAlaArg 59
Db      634 CAGTTGCTGGCGAATCGGTGACGATGCCCGCGCGAAGCCTTGACAAAGCCGCAAG 693
QY      60 ArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGlu 79
Db      694 CTGATCGGCGCTG---GGCTATCCC-----GGTGTGTCGGAAATCGCC 732
QY      80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAla 99
Db      733 CGCTGGCGGAGCGCGGACACTCCCTGCGCTTCGTTCCCGCGGCCGATGACCGATGCC 792
QY      100 LysAsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119
Db      799 CCGCGCTGAGACTTCAGCTTCAGCGGGCTCAAGACCTTACCTTAAC---ACCTGGCAG 849
QY      120 LysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 139
Db      850 CGTTGCGTGGAGCGCGGACGACGACGAGCAG-----ACCGCTGGCAGATCGCC 900
QY      140 AlaThrValGlnHisThrMetAlaCysHisIleValLysArgThrHisArgAlaIleLeu 159
Db      901 CTGGCGTTCCAGACCGCGGCTGTGAGACCTGCTGATCAAGTGCCTGCGCCCTTG--- 957
QY      160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyGly 179
Db      958 -----AAGCAGACCGGCTG-----AAGAAC-----CTGTATATCGCGCGGT 996
QY      180 ValAlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCys 199
Db      997 GTACGCCCAACAGAGCGCTGCGCAGCGGCTGGAAGAGATGCTCGCGAATGAAGAGGG 1056
```



APPLICATION NUMBER: 08/487,429  
FILING DATE: 1995-06-07  
APPLICATION NUMBER: 08/426,787  
FILING DATE: 1995-04-21  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PH186P1C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-610-5790  
TELEFAX: 310-309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-643-990A-1

Alignment Scores:  
Pred. No.: 1,68e-23 Length: 1830121  
Score: 289.50 Matches: 82  
Percent Similarity: 48.70% Conservative: 30  
Best Local Similarity: 35.65% Mismatches: 95  
Query Match: 20.90% Indels: 23  
Gaps: 7

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-09-643-990A-1 (1-1830121)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeu--ThrAsnLysValGluPheProPheLeu 19  
Db 552881 ATGAGAGGCAATTACTTGCCGCAATGCTGATGACAAATTCACCGCACTTCTTTGTT 552822  
QY 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspHe 39  
Db 552821 GCTTGTTGGTATCCGGTGGCCACCTCAATTATGCGCTGTGAGTGTGTAGAAAATAT 552762  
QY 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59  
Db 552761 GAAGTATAGAGAAATCTATTGATGATGCTGCTGCGAAAGCCTTTGATTAACAGCAAAA 552702  
QY 60 ArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGlu 79  
Db 552701 TTACTTGACATG--GATTATCCA-----GTTGGCCGGCACTTCT 552663  
QY 80 HisLeuAlaLysGlnGlyAsnArgPheHisAspAspIleLysProProLeuHisIleAla 99  
Db 552662 CGTTTAGCGGAAAAAGGTACGCCCAAAATGCTTCCACATTTCCAGTCAATGACAGATCGT 552603  
QY 100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119  
Db 552602 GCAGCGCTTGATTTAGTTTCTTGTAAACAAATTCGCCCAATGACATTAATCA 552543  
QY 120 LysGluLysGlnGluGly-----IleGluLysGlnIleLeuSerAlaIleAsp 137  
Db 552542 GCATTATAAAACGAGGCGCAACTGATAGCA-----ACTAAAGCAAGT 552498  
QY 138 IleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAla 157  
Db 552497 ATTCCTTATGCTTTCAGAGAGCGGTGTGATACCTTTGCC----- 552456  
QY 158 IleLeuPheCysLysGlnArgAspLeuProGlnAsnAsnAlaValLeuValAlaSer 177  
Db 552455 ---ATTAAATGTAG--CGTGCATTGAAGAAACAGGCTATAAACGTTTATGATTGG 552402  
QY 178 GlyValValAlaSerAspPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThr 197  
Db 552401 GGAAGGCGGAGCGCAAAATAAACTCCGAGAAACGCTTCCGCACTTAAGCAAAATTTA 552342  
QY 198 GlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAla 217  
Db 552341 GGTGGCGAAGTGTTTATCTCAACCTCAATTTTGTACAGATTAATGTCGCGATGTTGCT 552282

QY 218 TrpAsnGlyIleGluArgLeuArgAlaGly 227  
Db 552281 TACACAGTGTTTTACGTTTAAAAACAAGCT 552252

RESULT 12

US-09-902-540-6612  
Sequence 6612, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT APPLICATION NUMBER: US/09/902,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 6612  
LENGTH: 996  
TYPE: DNA  
ORGANISM: Myxococcus xanthus  
US-09-902-540-6612

Alignment Scores:

Pred. No.: 6,44e-28 Length: 996  
Score: 282.00 Matches: 86  
Percent Similarity: 48.93% Conservative: 28  
Best Local Similarity: 36.91% Mismatches: 87  
Query Match: 20.36% Indels: 32  
Gaps: 9

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-09-902-540-6612 (1-996)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysVal--GluPheProPheLeu 19  
Db 334 CTGAGAGGCACTCTGTGCGCATCCGGCTTTGAGAGTGGCCCGGAGCCCGCTTCTT 393  
QY 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspHe 39  
Db 394 GGGCTGCTGTTTCCGGCGGACACACAGCCTCTACAGAGTGTGACAGGCTTACCGGAGTAC 453  
QY 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59  
Db 454 CGGCTGTGGGAGCAGCGCGGACGAGCGCGCGGAGCATATGACAAAGCCGCTCCG 513  
QY 60 ArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGlu 79  
Db 514 ATCTCGGCGCTG--CCGATCCG-----GTTGGAGGCCCATGAC 552  
QY 80 HisLeuAlaLysGlnGlyAsn-----ArgPheHisAspIleLysProPro 95  
Db 553 CAGTTGCGCCACAGGCGAAACCCGAGGCAATCCGCTT-----CCGCC 597  
QY 96 LeuHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLys 115  
Db 598 GCGTGCAGGAGCAAACTTGCAGCGTCTCTCCGGGTTGAG-----ACG 645  
QY 116 IleIleMetLysLysGlnGlyGlnGluGlyIleGluLysGlyGlnIleLeuSerAla 135  
Db 646 GCGGCTGCACACCTGTGAGAAAGCGCGTGCAGGCGGAGCGGCGGCTG----- 696  
QY 136 AlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHis 155  
Db 697 GCGAATTTGTGCGCTTCTTCCAGAGGCGGTGGCGGAAGCTGTGTGAGAAAG----- 750  
QY 156 ArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuVal 175  
Db 751 -----CTGTGGCGCGCGCGCGCGGCTTG-----GCCACAAAGCAATTGGTG 792

QY 176 ALaSerGlyGlyValAlaSerAspNheThyrlleArgAlaLeuGluIleLeuThrAsn 195  
| | | | | : : : : :  
DB 793 CTGTGGCGCGCGCTGCGCGGAACTCGCGGCACTGTCTACAGCGGAGCGGAG 852  
QY 196 AlaThrGlnCysThrLeuLeuCysProProAlaGluCysThrAspAsnGlyIleMet 215  
| | | | | : : : : :  
DB 853 GAGCGGGGGTGAACATGTTCTCTGCCCCCGGCTGTGCAAGCAATGGCCGCAAG 912  
QY 216 lLeaIaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 228  
| | | | | : : : : :  
DB 913 ATTGGGTGGCGGGGTATGAGCGCTACCGCGCGGCTGTG 951

RESULT 13  
US-09-902-540-503/c  
; Sequence 503, Application US/0902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(115849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 503  
; LENGTH: 2582  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-503

Alignment Scores:  
Pred. No.: 3,15e-27 Length: 2582  
Score: 282.00 Matches: 86  
Percent Similarity: 48.93% Conservative: 28  
Best Local Similarity: 36.91% Mismatches: 87  
Query Match: 20.36% Indels: 32  
DB: 4 Gaps: 9

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-09-902-540-503 (1-2582)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIleVal---GluPheProPheLeu 19  
| | | | | : : : : :  
DB 1195 CTGAGGGGCAACCTGTGGCCATCCGGCTGTGAGGTGCGCCGAGCGCGGCTTCCTT 1136  
QY 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPhe 39  
| | | | | : : : : :  
DB 1135 GGGCTCGTCTTTCGGCGGGGACACAGCCCTTACAGAGTGCAGGCTTACGGGCACTAC 1076  
QY 40 LeuLeuLeuGlyIlySerLeuAspIleAlaProGlyAspMetLeuAspIlyValAlaArg 59  
| | | | | : : : : :  
DB 1075 CGGCTGTGGGAGACGCGGACGACGCGCGGAGGAGCATATGACAAGACCGCTCCG 1016  
QY 60 ArgLeuSerLeuIleIlyHisIleProGluCysSerThrMetSerGlyGlyIlyAlaIleGlu 79  
| | | | | : : : : :  
DB 1015 ATCTCGGCTG---CCGATATCCG-----GGTGGGAGCGCCATCGAC 977  
QY 80 HisLeuAlaIlyGlnGlyAsn-----ArgPheHisPheAspIleIlyAspProPro 95  
| | | | | : : : : :  
DB 976 CAGTGGGCGACAGAGGAAACCCGAGGCCATCCGCTC-----CCGCC 932  
QY 96 LeuHisAlaIlyAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIly 115  
| | | | | : : : : :  
DB 931 GCGCTCGCGGAGACACTTCGACGTCTCTCCGGCTTGAAG-----AGC 884  
QY 116 IleIleMetIlyIlyGlnGlyGlnGluGlyIleGlyIlyGlnIleLeuSerSerAla 135  
| | | | | : : : : :  
DB 883 GCGGTGTGCACCACTGTGAGAAAGACGCGGTGCGGAGGCGGAGCGCTG----- 833  
QY 136 AlaAspIleAlaIleThrValGlnHisIleThrMetAlaCysHisLeuValIlyArgThrHis 155

DB 832 GCGAATTTGTGGCGCTTTCACAGAGCGCGGAGCGGAGCTGTGCAAGAG----- 779  
QY 156 ArgAlaIleLeuPheCysIlyGlnArgAspLeuProGlnAsnAsnAlaValLeuVal 175  
| | | | | : : : : :  
DB 778 -----CTGTGGCGCGCGCGCGCGCTG-----GGCCACAAGACGTTGGTG 737  
QY 176 ALaSerGlyGlyValAlaSerAspNheThyrlleArgAlaLeuGluIleLeuThrAsn 195  
| | | | | : : : : :  
DB 736 CTGTGGCGCGGCTGCGCGGAACTCGCGGCTGCGGCACTGTCTACAGCGGAGCGGAG 677  
QY 196 AlaThrGlnCysThrLeuLeuCysProProAlaGluCysThrAspAsnGlyIleMet 215  
| | | | | : : : : :  
DB 676 GAGCGGGGGTGAACATGTTCTCTGCCCCCGGCTGTGCAAGCAATGGCCGCAAG 617  
QY 216 lLeaIaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 228  
| | | | | : : : : :  
DB 616 ATTGGGTGGCGGGGTATGAGCGCTACCGCGCGGCTGTG 578

RESULT 14  
US-09-543-681A-2341  
; Sequence 2341, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 2341  
; LENGTH: 1074  
; TYPE: DNA  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-2341

Alignment Scores:  
Pred. No.: 1,17e-27 Length: 1074  
Score: 280.50 Matches: 81  
Percent Similarity: 46.77% Conservative: 35  
Best Local Similarity: 32.66% Mismatches: 101  
Query Match: 20.25% Indels: 31  
DB: 4 Gaps: 8

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-09-543-681A-2341 (1-1074)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIleVal---GluPheProPheLeu 19  
| | | | | : : : : :  
DB 385 ATGAAAGGCAATTTATTAAGCGCGGATGCTTGAAGAAAGAACCCAGATTTCTTTCGTG 444  
QY 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPhe 39  
| | | | | : : : : :  
DB 445 GCGTACTTGTCTCGGGGGGAGCATACACAATTAATTAAGTGAACAGAGATGGGAATAT 504  
QY 40 LeuLeuLeuGlyIlySerLeuAspIleAlaProGlyAspMetLeuAspIlyValAlaArg 59  
| | | | | : : : : :  
DB 505 ACCCTGTAGTGACGTATGATGATGCTGTGCTGAGACATTTGATTAAGAGCCAG 564  
QY 60 ArgLeuSerLeuIleIlyHisIleProGluCysSerThrMetSerGlyGlyIlyAlaIleGlu 79  
| | | | | : : : : :  
DB 565 CTATTGGGCTT---GATTATCC-----GGCGGCTGTATTATCA 603  
QY 80 HisLeuAlaIlyGlnGlyAsnArgPheHisPheAspIleIlyAspProProLeuHisAla 99  
| | | | | : : : : :  
DB 604 AAAATGGCACACAGGTGTGAAGAGACGTTTGTCTTCTCGTCCAGACAGACAGA 663  
QY 100 IlyAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIlyIleIleMetIly 119  
| | | | | : : : : :  
DB 664 CCGGAGCTGCACCTTATGTTCTGTGTTAAACCTTTGCGCTAATCTATTTCTGTCA 723

```

Qy 120 LysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaIleAla 139
Db 724 AACCATGATTGACAG-----CAAATCGACGACATATTGCG 759
Qy 140 AlathrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeu 159
Db 760 CGTCTTTTGAAGATCCGCTGATAGTACTTTGGCAATAAATGTCGTGACCA----- 813
Qy 160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsn---AlaValLeuValAlaSerGly 178
Db 814 -----TTAGACCAACAGCCTTTAAACGCTTGATGATGCTGG 852
Qy 179 GlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGln 198
Db 853 GGCCTTAGTGTAAACCGTACTTACGCCCAAAATGCGCATATATGACAACTCGGA 912
Qy 199 CysThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrp 218
Db 913 GGGGAAAGTTTATGCTGCGCTGAGTTATGTACCGAATATGTCCTGATGCTTGG 972
Qy 219 AsnGlyIleGluArgLeuArgAlaGlyLeu---GlyIleLeuHisAspIleGluGlyIle 237
Db 973 GCGGGAGATGATCCGTTTAAAGTGTACCGACGCGCTTTA-----GGCGTG 1020
Qy 238 ArgTyrGluProLysCysProLeu 245
Db 1021 ACAGTGAGACCACTGGCTTTA 1044

```

```

RESULT 15
US-08-087-797-1
; Sequence 1, Application US/08087797
; Patent No. 5543112
; GENERAL INFORMATION:
; APPLICANT: Mellors, Alan
; APPLICANT: Lo, Reggie Y.C.
; APPLICANT: Abdullah, Khalid M.
; TITLE OF INVENTION: Pasteurella Haemolytica
; TITLE OF INVENTION: Glycoproteinase
; TITLE OF INVENTION: Gene and the Purified Enzyme
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.
; STREET: 1211 East Morehead Street,
; CITY: Charlotte
; STATE: No. 5543112th Carolina
; COUNTRY: United States
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/087,797
; FILING DATE: 14-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Layton, Jr., Samuel G.
; REGISTRATION NUMBER: 22807
; REFERENCE/DOCKET NUMBER: 3374-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 704 377 1561
; TELEFAX: 704 334 2014
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1315 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: linear
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEtical: NO
; ANTI-SENSE: NO
; FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: 141..1115
; US-08-087-797-1
Alignment Scores:
Pred. No.: 3,72e-26 Length: 1315
Score: 270.50 Matches: 79
Percent Similarity: 46.46 Conservative: 26
Best Local Similarity: 34.96 Mismatches: 102
Query Match: 19.53 Indels: 19
DB: 1 Gaps: 6
US-10-649-273-2_copy_148_414 (1-267) x US-08-087-797-1 (1-1315)
Qy 1 MetGluAlaHisAlaLeuThrIleArgLeu---ThrAsnLysValGluPheProPheLeu 19
Db 474 ATGGAAGGCAATTACTTCCCAATGTTGGAAGAAATGCCCTGATTTGCTTGG 533
Qy 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPhe 39
Db 534 GCATTTATGATTTCAAGTGGACACACCAACTGTGTAAGTTGACCGCGTTGGCAATAC 593
Qy 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59
Db 594 GAACCTAGCGGGAATCAATTATGATGCTGCCGCGTGAAGCCTTTGACAAACAGGCAGA 653
Qy 60 ArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyValLysAlaIleGlu 79
Db 654 CTACTCGCTTG---GATTACCT-----GCCGCTGACGATGCTCA 692
Qy 80 HisLeuAlaLysGlnGlyAsnArgPheHisAspAspIleLysProLeuHisAla 99
Db 693 AATATGACCAATCCGACGACCAATGCTTTAAATTCCTGTCATGACCGACAGA 752
Qy 100 LysAsnCysAspPheSerThrThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119
Db 753 CCGGAGCTGATTTCAAGTTCCTCGGTTAAACCTTTGCTCGCAATACGATTAAGCC 812
Qy 120 LysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaIleAla 139
Db 813 AATCTTAATGAATAATGTAAGTCAATGACGAA-----ACCAATGCCATATTGCC 863
Qy 140 AlathrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeu 159
Db 864 CACGCAATCCAAACAGCCGTC-----GTTGATCTATTTTAATT 902
Qy 160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyIle 179
Db 903 AATGCAAG---CGAGCGTTAGAGCAACCGGCTAATTAACGCTTAATGATGCGACGCGC 959
Qy 180 ValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCys 199
Db 960 GTTAGGCAATTAACAAATTAACGACGACCTTGGCAATATGTAAGAAATTAAGGC 1019
Qy 200 ThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsn 219
Db 1020 GAAGTATTTACTTCCCTGCCACAAATTTTGCATGACACGCGCAATGATTTGCTTACT 1079
Qy 220 GlyIleGluArgLeuArg 225
Db 1080 GCGCTTCTTCCGCTTAAA 1097

```

Search completed: November 11, 2005, 03:03:04  
Job time : 1105.51 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 10, 2005, 16:53:13 ; Search time 3784.43 Seconds  
(without alignment)  
2685.516 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414

Perfect score: 1385  
Sequence: 1 MEAHALITRLTNKVFPPFLV.....DISKVGASIKVQLKMEI 267

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Zgapop 6.0 , Zgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp  
-O=/cgp2.1/USPFO\_spool\_p/USY0649273/runat\_02112005\_091338\_15562/app\_query.fasta\_1.1429  
-DB=EST -QPM=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human4.0.cdi -LIST=45  
-DOCFALGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273 @CGN 1 1 8076 @runat 02112005\_091338\_15562 -NCPU=6 -ICPU=3  
-NO\_MMAP -IARBOURRY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hic: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_ges1: \*  
9: gb\_ges2: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1263	91.2	870	5	BQ423651 AGENCOURT
2	1224	88.4	2284	3	AK045669 Mus muscu
3	1216	87.8	1622	3	AK011265 Mus muscu
4	1097	79.2	640	5	B0636028 hnd03d11.y
5	1075	77.6	852	3	BX391919 BX391919
6	1071	77.3	1563	3	BCU30671 Mus muscu
7	1053	76.0	701	2	BE740611 B0961028
8	944	68.2	922	5	B0961028 AGENCOURT
9	925	66.8	658	7	CF362328 829596 MA

10	916	66.1	822	7	CF257246
11	838	60.5	637	7	CK941819
12	830	59.9	597	6	CB272391 ma157912.
13	819.5	59.2	792	5	BU403563
14	816	58.9	730	7	CN823245
15	789	57.8	545	1	AV602901
16	786	56.0	866	5	BU127463
17	784	56.6	490	6	CB852881
18	749	54.1	736	7	CK365185
19	749	54.1	1173	6	CD508917
20	746	53.9	723	5	BU261251
21	722	52.1	812	5	BU264489
22	717	51.8	634	2	AM601179
23	709	51.2	484	1	AJ670918
24	708	51.1	736	6	CA057753
25	707.5	50.1	701	5	BU621780
26	704	50.8	749	7	CK982692
27	698	50.4	696	5	BU102666
28	692	50.0	579	9	CK819035
29	691.5	49.9	789	5	BU242187
30	683	49.3	909	5	BX756548
31	682	49.2	548	7	CO880741
32	671	48.4	682	2	BB043703
33	668	48.2	706	5	BU202465
34	666	48.1	869	5	BX754527
35	647	46.7	919	7	CF407294
36	646	46.6	878	5	BX776940
37	645	46.6	533	4	BM126453
38	638	46.1	1082	5	BX359023
39	625	45.1	506	2	BF415802
40	618.5	44.7	879	5	BU256052
41	614	44.3	861	5	BU246158
42	611	44.1	1171	5	BU261605
43	604	43.6	424	1	AA273921
44	602	43.5	413	1	AA589724
45	600	43.3	863	5	BU376295

#### ALIGNMENTS

RESULT 1  
BQ423651 870 bp mRNA linear EST 23-MAY-2002  
LOCUS BQ423651  
DEFINITION AGENCOURT 7790948 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:605828  
5', mRNA sequence.

ACCESSION BQ423651.1 GI:21118966  
VERSION BQ423651.1  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 870)

AUTHORS NIH-MGC http://mhc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov

Tissue Procurement: ATCC/DCFD/DP  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Place: LLM13342 row: 1 column: 21  
High quality sequence stop: 710.

#### FEATURES

source  
1..870  
location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:605828"





Genomic Sciences Central and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: <http://genome.gsc.riken.jp/>  
URL: <http://fantom.gsc.riken.jp/>.  
Location/Qualifiers

## FEATURES

SOURCE

1..2284  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:B330219017"  
/db\_xref="taxon:10090"  
/clone="B230219017"  
/sex="male"  
/tissue\_type="corpora quadrigemina"  
/clone\_id="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
221..1465  
/note="unamed protein product; putative  
similar to PUTATIVE SIALOGLYCOPROTEASE TYPE 2 [Homo  
sapiens] (SPTR|O9H4B0, evidence: FASTY, 80%ID, 100%length,  
match=1242) "  
/codon\_start=1  
/protein\_id="BAC32450.1"  
/db\_xref="GI:26337529"  
/translation="MLMLRRTAGAIKPKPKSVYGFLLRFRVHPRTLCHKLVLIET  
SCDPTGAIVDEITVGEALHSGTOVALTKGIVPAQOOLRENIQIVETLSAC  
RTSPDSALATTKIKGLALSLGVLSFSLDLOVFPKPRPIHMEBALITRLTNK  
VEPPLVILISGHCFLALVGVSDPLILGSLDIPADMDLKVARIJSLIHPREST  
MSGKALIEQLAKDGRFHFTLNPQNAKNCDFSTGLQHTIDKLIITHEKEGIEKG  
QILSSADIAAIVOHATKLAIRKTRHALIFPCQKRLSPANAIVVSGVSNLYIR  
KALEIVANATCCTLCPPRLCTDNGIMIAMNGIERTLRAGLGLHVEDIRYEPKPL  
GVDISREVAEAIKVRKML"

CDS

polyA\_signal  
2262..2267  
/note="putative"  
2284  
polyA\_site  
/note="putative"  
ORIGIN

Alignment Scores:  
Pred. No.: 2.75e-131 Length: 2284  
Score: 1224.00 Matches: 23  
Percent Similarity: 92.51% Conservative: 14  
Best Local Similarity: 87.27% Mismatches: 20  
Query Match: 88.38% Indels: 0  
Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x AK045669 (1-2284)

QY 1 MetGluAhiSaIaLeuThrIleArgLeuThraenLysValGluPheProPheLeuVal 20  
Db 662 ATGGAGGCTCAGCAGTCACTATTAGGCTCACCAATTAAGATATTCCTTTTATGTT 721  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGInGlyValSerAppheLeu 40  
Db 722 CTTTGTGATTTTGGCGGTCACTGCTGTGGCATTAGTCCAAAGTGTTTCCATTCCG 781  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
Db 782 CTCCTGGAGAGTCTTTCGACATAGCACAGCGACATGCTTGACAAAGTGGCAAGAGA 841  
QY 61 LeuSerLeuIleLysHisProGlyLysSerThrMetSerGlyGlyLysAlaIleGluHis 80  
Db 842 CTTTCTTTAATCAACATCCAGAAAGTCTCAAGAGTGGGAAAGCTATTAGAAACAG 901  
QY 81 LeuAlaLysGInGlyAsnArgPheHisPheAspIleLysProProLeuHisIleAlaLys 100  
Db 902 TTGGCCAAAGCGAAATAGATTCCATTCTACTATCAATCCACTATGACAGATCTAAG 961  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValIthraSpLysIleIleMetLysLys 120  
Db 962 AATTGGCAATTTTCTTTCACGGGACTTCAACATATTACTGATTAAGCTAATAACACAAAG 1021

QY 121 GluLysGluGluGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaSerIleAlaAla 140  
Db 1022 GAAAAAGAAAGAGGCAATTCAGAGGGCAAAATCCTGTCATCACTGCAGACATTCGCT 1081  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
Db 1082 GCGGTACAGCATGCAACAGGTGCCACTTGGCAAAAAGAACACATCCGCGTATTCGTTT 1141  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
Db 1142 TGGACACGAAAAATTTGCTCTCTCTCCAGCTACACGCAATTAAGTTGATTCGAGGTGT 1201  
QY 181 AlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
Db 1202 GCAAGTACCTTGATACATCCGAAAGACATTGGAATTTGGCAATGCAACGACGTGACG 1261  
QY 201 LeuLeuCyProProProArgLeuCyThrAspAsnGlyIleMetIleAlaTPanGly 220  
Db 1262 TTGTTGTCTCCACCTCCAGACACTGTGCACCTGACCAATGCAATTCATGATGGAATGA 1321  
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlyGlyIleArgTyrglu 240  
Db 1322 ATTTGAAGATTACGTCTGCTGGCGTTTACATGATGTAGAGACATCCATGATGAA 1381  
QY 241 ProLysCyProLeuGlyValAspIleSerLysGlyValGlyGluAlaSerIleLysVal 260  
Db 1382 CCAAAATGCTCTTTCGAGTGAAGCATATCCAGAGAAATTCACAGACGTCCATTAAGTA 1441  
QY 261 ProGlnLeuLysMetGluIle 267  
Db 1442 CCGCATTAATAAATGCACATT 1462

## RESULT 3

AK011265 1622 bp mRNA linear HTC 03-APR-2004  
LOCUS Mus musculus 10 days embryo whole body cDNA, RIKEN full-length  
DEFINITION enriched library, clone:261001M19 product:similar to PUTATIVE  
SIALOGLYCOPROTEASE TYPE 2 [Homo sapiens], full insert sequence.

ACCESSION AK011265 GI:12847275  
VERSION AK011265.1 GI:12847275  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE 1  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
AUTHORS Komno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636

TITLE 2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL MEDLINE 2049374  
PUBMED 11042159  
REFERENCE 3  
AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,  
Komno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--394-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
JOURNAL MEDLINE 20530913  
PUBMED 11076861



BO636028  
LOCUS B0636028 640 bp mRNA linear EST 15-UTL-2002  
DEFINITION h0d03d11.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he  
ACCESSION Homo sapiens cDNA clone h0d03d11.5', mRNA sequence.  
VERSION B0636028  
KEYWORDS B0636028.1 GI:21760487  
SOURCE EST.  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 640)  
REFERENCE Wistow, G., Bernstein, S.L., Wyatt, M.K., Ray, S., Behal, A.,  
Touchman, J.W., Bouffard, G., Smith, D. and Peterson, K.  
Expressed sequence tag analysis of human retina for the NEIBank  
Project: Retbindin, an abundant, novel retinal cDNA and alternative  
splicing of other retina-preferred gene transcripts  
Mol. Vis. 8 (4), 196-204 (2002)  
JOURNAL  
MEDLINE  
PubMed  
COMMENT Contact: Wistow G  
Section on Molecular Structure and Function  
National Eye Institute  
6/331, NIH, Bethesda, MD 20892-2740, USA  
Tel: 301 402 3452  
Fax: 301 496 0078  
Email: graeme@helix.nih.gov  
Plate: 03 row: d column: 11  
Seq primer: M13RP1 reverse primer (ABI).  
Location/Qualifiers  
1. 640  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="h0d03d11"  
/issue\_type="Retina"  
/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
/clone\_lib="Human Retina cDNA (Un-normalized,  
unamplified): hd/he"  
/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue  
was dissected from two 80 year old donors with no observed  
eye disease. 100ng of total RNA was used for library  
construction. A directionally cloned cDNA library in the  
pSPORT1 vector (Life Technologies) was constructed at  
Bioserve Biotechnology (laurel MD) essentially following  
the protocols of the SuperScript Plasmid System full  
details of which are contained in the manufacturer's  
instruction manual (http://www.lifetech.com/). First  
strand synthesis was carried out using a Not I  
primer-adaptor  
[5'-pGACTGCTTCTAGATCGCGAGCGCCGCC(T)15-3']. EST analysis  
was performed on the unamplified library at the NIH  
Intramural Sequencing Center (NISC)."  
ORIGIN  
Alignment Scores:  
Pred. No.: 2.94e-117 Length: 640  
Score: 1097.00 Matches: 209  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 79.21% Indels: 0  
DB: Gaps: 0  
US-10-649-273-2\_copy\_148\_414 (1-267) x B0636028 (1-640)  
Qy 59 AATGATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 78  
Db 3 AGAAGACTTCTTAAATTAACATTCAGAGGCTCCACACATAGATGAGGAGAAACCAATA 62  
Qy 79 GATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 98  
Db 63 GATGCTTGGCCAAACAGAAATGATTTTCACTTTGACATCAACCTCCCTGCATCAT 122

Qy 99 AATGATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 118  
Db 123 GCTAAAGAAATGCTATTTCTTTTACTGACCTTCAACACGTTATGATAAATATATATG 182  
Qy 119 TATGATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 138  
Db 183 AATGATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 242  
Qy 139 AATGATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 158  
Db 243 GCTGCTGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 302  
Qy 159 TATGATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 178  
Db 303 GCTGCTGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 362  
Qy 179 TATGATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 198  
Db 363 GCTGCTGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 422  
Qy 199 TATGATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 218  
Db 423 TATGATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 482  
Qy 219 TATGATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 238  
Db 483 TATGATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 542  
Qy 239 TATGATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 258  
Db 543 TATGATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 602  
Qy 259 TATGATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 267  
Db 603 TATGATGCTGCTTCTAGATCGCGAGCGCCGCC(T)15-3' 629  
RESULT 5  
LOCUS BX391919/c 852 bp. mRNA linear EST 28-APR-2004  
DEFINITION BX391919 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens  
cDNA clone CSDBK001Y802 3-PRIME, mRNA sequence.  
ACCESSION BX391919  
VERSION BX391919.2 GI:46846154  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 852)  
REFERENCE Li, W.B., Gruber, C., Jessee, J. and Polyes, D.  
Full-length cDNA libraries and normalization.  
Unpublished (2001)  
On May 13, 2003 this sequence version replaced gi:30611736.  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
1240.x  
For more information about this cluster, see  
http://www.genoscope.cns.fr/cdna?cs=CSDBA103552E07\_CS03317\_1&c=1240.x  
FEATURES  
source Location/Qualifiers  
1..852  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSDBK001Y802"

```
/cell_type="HELA CELLS COT 25-NORMALIZED"
/clone_lib="HELA"
/notes="First strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
```

## ORIGIN

## Alignment Scores:

Pred. No.:	1.65e-114	Length:	852
Score:	1075.00	Matches:	208
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	77.62%	Indels:	0
DB:	5	Gaps:	0

US-10-649-273-2\_COPY\_148\_414 (1-267) x BK391919 (1-852)

```
Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAenLysValGluPheProPheLeuVal 20
Db 626 ATGAGAGCTCATGCATCTTACCTTATAGTTGACCATTAAGTGAATTCCTTTTACGT 567
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 566 CTTTGTGATTTCTGGAGGTCACGCTCTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCTG 507
Qy 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db 506 CTTTGTGAAAGCTCTTGGACATAGCACAGGTGACAGCTTGACAAAGTGGCAAGAGA 447
Qy 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGlnHis 80
Db 446 CTTTCTTTAATTAACATCCAGAGTCTCCACCATGATGATGGTGGGAAAGCCATTAGACAT 387
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisAspIleLysProProLeuHisHisAlaLys 100
Db 386 TTGGCCAAACAAAGAAATAGATTCATTTTGAATCAACACCTCTTGATCATGCTAA 327
Qy 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db 326 AATTGTGATTTTCTTTTACCTGACCTTCAACCGCTTACTGATTAATATATGAAAAAG 267
Qy 121 GluLysGlnGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIa 140
Db 266 GAAAGAGAGAGGATATTGAGAGAGGCAATCTGTCTTCCAGCAGACATTTGCTGCC 207
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
Db 206 ACAGTACAGCACACATGCGATGTCATCTTGGAAAGAACACATCGGCTATTCGTTT 147
Qy 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180
Db 146 TGTAAAGCAGAGAGACTTGTACTTCAAAATATATGACATGCTGTGTCATCTGTGTGTC 87
Qy 181 AlaSerAsnPhePheTrpIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
Db 86 GCAAGTAACTTCTATATCCGAGAGCTCTGAAATTTTAAACAAACGCAACACAGTGCACT 27
Qy 201 LeuLeuCysProProArgLeu 208
Db 26 TTGTTGTCTCTCTCTCCAGACTA 3
```

```
RESULT 6
LOCUS BC030671 1609 bp mRNA linear HTC 19-NOV-2003
DEFINITION Mus musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (cDNA
clone IMAGE:1226118), containing frame-shift errors.
ACCESSION BC030671
VERSION BC030671.1 GI:21040459
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
```

REFERENCE  
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1609)

Klausner, R.D., Collins, F.S., Wagner, L.H., Derge, J.G., Strassberg, R.L., Feingold, R.A., Grouse, L.H., Schuler, G.D., Altschul, S.F., Zeeberg, B., Burow, K.H., Scheffer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Wax, S.I., Wang, J., Heish, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mulhally, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huiyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butcherfield, Y.S., Krzywinski, M.I., Skalska, U., Smallus, D.E., Scherch, A., Schein, J.R., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 1609)

Strassberg, R.

Direct Submission

Submitted (20-May-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK  
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

Tissue Procurement: Marcello Bento Soares, Ph.D.

cDNA Library Preparation: M. Bento Soares, University of Iowa

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

contact: [amadan@systemsbiology.org](mailto:amadan@systemsbiology.org)

Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLML at: <http://image.llnl.gov>

Series: IRK Plate: 66 Row: e Column: 10

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21312463

This clone has the following problem: frame shifted.

## FEATURES

source

1..1609

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:1226118"

/issue\_type="Thymus gland, mouse"

/clone\_lib="Soares\_thymus\_2N2MT"

/lab\_host="DH10B"

/note="Vector: pT73-Pac"

## ORIGIN

Alignment Scores:

Pred. No.:	1.21e-113	Length:	1609
Score:	1071.00	Matches:	211
Percent Similarity:	83.52%	Conservative:	12
Best Local Similarity:	79.03%	Mismatches:	20
Query Match:	77.33%	Indels:	24
DB:	3	Gaps:	1

US-10-649-273-2\_COPY\_148\_414 (1-267) x BC030671 (1-1609)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
Db 665 ATGAGAGGCTCAGCAGCTACTGATTAGGCTCACCAGTAAGTAGAATTCCTTTTATGTT 724  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
Db 725 CTTTGATTTCTGGGGCTACGCTGTTGGCATTAGTCCAAAGGTGTTCCATTTCCCTG 784  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
Db 785 CTCCTTGGAGAGCTTTGGACATACGACACGCGACATCTTGACAGGTGGCAAGAGA 844  
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysAlaIleGluHis 80  
Db 845 CTTTCTTAATCAAAACATCCAGATGTTCTCAATGAGTGCGGAAAGACTATAGAACAG 904  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
Db 905 TTGGCCAAAGACGGAAATAGATTCCATTACTCAATCCACCTATGCAAGATGCTAAG 964  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
Db 965 AATTGCCATTTTCTTTACCGGACTTCAACATATTACTGATTAAGCTAATACACACAG 1024  
QY 121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
Db 1025 GAAAAGAGAAAGGACATTGAGAGAGGGCAATCTGTATCACTGACGACATTTGCTGT 1084  
QY 141 ThrValGlnHisIsthMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
Db 1085 GCGGTACAGCATGCAACACCGCGCACCTTGCAGAAAGAACACATCGCGCTTCTGTTT 1144  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
Db 1145 TGCAGGACGAAAAATTTCTCTCTCCAGCTAACGAGATTAAGTTGATCTGAGAGTGT 1204  
QY 181 AlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
Db 1205 GCAAGTAACTTGATCATCCGAAAGCATTTGGAAATTTGCGCAATGCAAGCGACGACG 1264  
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
Db 1265 TTGTTGTGTCCACCTCCAAAGACTGTGCACTGCAATGGCATCATGATTGCA----- 1315  
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrlu 240  
Db 1315 ----- 1315  
QY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyLysValAlaSerIleLysVal 260  
Db 1316 ---TGATGTCCTCTTGGAGTACATATCCAGAGAGGTTCGAAAGCTGCCATTAAGTA 1372  
QY 261 ProGlnLeuLysMetGluIle 267  
Db 1373 CCGGCAATTAAAAATGCACTT 1393  
RESULT 7  
BE740611 701 bp mRNA linear EST 15-SEP-2000  
LOCUS 601595739P1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3949640 5',  
DEFINITION mRNA sequence.  
ACCESSION BE740611 GI:10154603  
VERSION BE740611.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
NIH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: DCTD/DRP  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at: [image.lnl.gov](http://image.lnl.gov)  
plate: LNCM814 row: n column: 09  
High quality sequence stop: 701.  
Location/Qualifiers  
1..701  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3949640"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_9"  
/note="Organ: ovary; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; CDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCAAGAG(C). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN  
Alignment Scores:  
Pred. No.: 4,6e-112 Length: 701  
Score: 1053.00 Matches: 206  
Percent Similarity: 99.05% Conservative: 3  
Best Local Similarity: 97.63% Mismatches: 0  
Query Match: 76.03% Indels: 2  
DB: 2 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x BE740611 (1-701)  
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
Db 71 ATGAGAGGCTCAGCAGCTACTGATTAGGCTCACCAGTAAGTAGAATTCCTTTTATGTT 130  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
Db 131 CTTTGATTTCTGGAGTACATGTCGTTGGCATTAGTCCAAAGGTGTTCCATTTCCCTG 190  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
Db 191 CTTTGGAGAGGCTTTTGGACATACGACACGAGGACATGCTTGACAGGTGGCAAGAGA 250  
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysAlaIleGluHis 80  
Db 251 CTTTCTTAATCAAAACATCCAGATGTTCTCAATGAGTGCGGAAAGACTATAGAACAT 310  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
Db 311 TTGGCCAAAGACGGAAATAGATTCCATTACTCAATCCACCTATGCAAGATGCTAAG 370  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
Db 371 AATTGCCATTTTCTTTACCGGACTTCAACATATTACTGATTAAGCTAATACACACAG 430  
QY 120 yGluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
Db 431 GAAAAGAGAGAGGATATGAGAGGGGCAATCTGCTTCAAGACGACATGCTG 490  
QY 140 IatThrValGlnHisIsthMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuP 160  
Db 491 CCACAGTACAGCACCAATGAGATGTCATCTTGTAAAGAAACACATCGGCTATTCCT 550  
QY 160 heCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180



Db 551 TTGTGAACGACAGACTGTTCCTCAAAATATACGACTGCTGTCATCTGTCGTG 610

Qy 180 a1a1aSerAspPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysT 200

Db 611 TCGCAAGTAACCTTATATCCGACAGACTCTGGAAATTTTAACAAACGACACAGTCA 670

Qy 200 hrLeuLeuCysProProProArgLeuCys 209

Db 671 CTTGTGTGTCTCTCTCCAGACTATGC 699

RESULT 8

LOCUS BQ961028 922 bp mRNA linear EST 21-AUG-2002

DEFINITION AGENCOURT 8863711 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:6423902

ACCESSION BQ961028

VERSION BQ961028.1 GI:22376506

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 922)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov).

Tissue Procurement: DCTD/DTF/Cazdar

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ULNL at: <http://image.liml.gov>

Plate: LNCM2603 row: d column: 15

High quality sequence stop: 584.

Location/Qualifiers

1..922

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6423902"

/tissue\_type="large cell carcinoma"

/lab\_host="RDH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_18"

/note="Organ: Lung; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN

Alignment Scores:

Pred. No: 3.65e-99 Length: 922

Score: 944.00 Matches: 207

Percent Similarity: 89.45% Conservative: 5

Best Local Similarity: 87.34% Mismatches: 16

Query Match: 68.16% Indels: 10

DB: 5 Gaps: 3

US-10-649-273-2\_COPY\_148\_414 (1-267) x BQ961028 (1-922)

Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20

Db 207 ATGGAAGGCTCARGCACTTACTATTAGGTGACCAATAAGATTTCTTTTATAGTT 266

Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40

Db 267 CTTTGAATTTCTGGAAGGTCACGTCTGTGGCATTAGTTCAAGAGATTTCACATTTTCTG 326

Qy 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60

Db 327 CTTCTGGAAAGACTCTTGGACATAGCACAGGAGCATCTTGCACAAGGGCGAAGAAG 386

Qy 61 LeuSerLeuIleLysHisProGluCys-SerThrMetSerGlyGlyValAlaIleGluHis 80

Db 387 CTTTCTTTAATATAACATCCAGAGTGCNTCCACCAAGTGTGGGAAAGCCATTAACCA 446

Qy 80 sLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100

Db 447 TTGGCCAAACAGAAATAGATTTCATTGATTCATCAACCTCCCTTGATCATATCTTA 506

Qy 100 sAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysIly 120

Db 507 AATATTGTATTTTCTTTTACTGACCTTCAACCGTTACTGATTAATATATATGAAAAA 566

Qy 120 sGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaI 140

Db 567 GGAAAAAGAGGAAGGATTTAGAGAAAGGCAAAATCCGTCTTCAGCAGCAGACATTCCTGC 626

Qy 140 aThrValGlnHisSThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuP 160

Db 627 CACAGTACAGCACACATGCGCATGTCACTTGTGAAAAAGAAA-CATCGGAGCTATTTCTGT 685

Qy 160 eCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSer-GlyGly 180

Db 686 TTGTAAAGCAGACAGACTTGTAACCTCAAAATATATGCAAGTACTGGTGCACTTGGGGGG 745

Qy 180 a1a1aSerAspPheTyrIleArgArgAlaLeuGluIleLeuThr-AsnAlaThrGlnCys 199

Db 746 TCGCAAGTAACCTTATATATCCGCAAACTCTGGAAAAATTTACAAACCCACACCGGGG 805

Qy 200 Thr-LeuLeuCysProProProArgLeuCysThr-AspAsnGlyIleMetIleAlaTrp 218

Db 806 ACCTTTGTGGGTCCCCCTCCAAACTATGCACTGGATATATGGG---CATTAATGATGGC 862

Qy 219 -----AsnGlyIleGluArg---LeuArgAlaGlyLeuGly 229

Db 863 TCGGGAAGGGAATTTGAAAAAATAATACCTGCTCGCTTGGGG 903

RESULT 9

CF362328/ c 658 bp mRNA linear EST 25-AUG-2003

LOCUS CF362328/ c 658 bp mRNA linear EST 25-AUG-2003

DEFINITION 829596 MARC 3P1G Sus scrofa cDNA 3', mRNA sequence.

ACCESSION CF362328

VERSION CF362328.1 GI:34161882

KEYWORDS EST.

SOURCE Sus scrofa (pig)

ORGANISM Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

AUTHORS 1 (bases 1 to 658)

Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J., Wise,T.A., Nonneman,D.J., Wray,J.E. and Keeler,J.W.

A second set of porcine ESTs from a pooled-tissue normalized library

TITLE Unpublished (2003)

JOURNAL Contact: Smith TPL

COMMENT USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: [smith@mail.marc.usda.gov](mailto:smith@mail.marc.usda.gov)

Single pass sequencing: Bases called with phred v0.020425.c and trimmed with the aid of the trim\_alt option. Vector identified with cross\_match v0.990329.

plate: SRG8015 row: H column: 12

Seq primer: TGAAGGCAAGTCGAGG.

FEATURES

source

1..658

/organism="Sus scrofa"

/mol\_type="mRNA"



/db\_xref="taxon:9823"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/clone\_id="MARC 3P1G"  
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;  
Library made with RNA pooled from multiple tissues  
including brain, liver, muscle, placenta/endometrium,  
ovary, testes, and bone marrow."

## ORIGIN

## Alignment Scores:

Pred. No.:	3 71e-97	Length:	658
Score:	925.00	Matches:	174
Percent Similarity:	93.97%	Conservative:	13
Best Local Similarity:	87.44%	Mismatches:	12
Query Match:	66.79%	Indels:	0
DB:	7	Gaps:	0

US-10-649-273-2\_COPY\_148\_414 (1-267) x CF362328 (1-658)

QY 20 ValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuValGlnGlyValSerAspPhe 39  
DB 657 GTTCTTTGATATCTGGCGGATCTGCTTTGGCATTAGTAGAGAGGTTTCAGATT 598  
QY 40 LeuLeuLeuGlyLeuSerLeuAspPheLeuValProGlyAspMetLeuAspLysValAlaArg 59  
DB 597 CTGCTCTTGACAGCTTTGGACATAGCACAGGTGACATGCTTGACAGGTGACAGTA 538  
QY 60 ArgLeuSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 79  
DB 537 AGACTTTCTTAAATTAACATCCAGAGTGTCCACCATGATGTGGGAAAGCCATTA 478  
QY 80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspPheLeuValProLeuHisAla 99  
DB 477 CATTGTCCAAACGGAATTAAGTTCATTGATTCCTCCATGCAAGCTAT 418  
QY 100 LysAsnCysAspPheSerPheThrgLysLeuGlnHisValThrAspLysLeuLeuLys 119  
DB 417 AAAAATTTGATTTTCTTTTCTTGACCTCAACATGTTATTAAGACAATATAGCAG 358  
QY 120 LysGlnLysGlnGlyLysLeuLysGlnGlnLeuSerSerAlaAlaAspPheAla 139  
DB 357 AAGGAAAGAGAGAGAGATTTGAGAGAGGCAATCTGCTTCAAGCTGACAGATTGCT 298  
QY 140 AlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaLeu 159  
DB 297 GCGCAGATACAGACACAGATGCTGCATATTCCTCAAAAGACCTCATGCTATTCG 238  
QY 160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGly 179  
DB 237 TTTTGCACACAGAGAGCTTATTAATGTCACAAAGTAATGCAATGTTGTTATCTGAGGT 178  
QY 180 ValAlaSerAsnPheThrgLysArgAlaLeuGlnLeuLeuThrAsnAlaThrGlnCys 199  
DB 177 GTTGCAAGTAATTAATATACGAAAGCTTTAGAAAGTTGAGCAAAATGCAACAATG 118  
QY 200 ThrLeuLeuCysProProArgLeuCysThrAspAsnGlyLysLeuMetLeuAlaTrp 218  
DB 117 ACTCTGTGTGCTCTCTCCAGCATATGACATGATATAGCATTAATGATTGATGG 61

## RESULT 10

CF257246 822 bp mRNA linear EST 07-AUG-2003

LOCUS pha008\_g02 PHA-activated splenocytes Gallus gallus cDNA, mRNA

## DEFINITION

CF257246 CF257246

VERSION CF257246.1 GI:33490501

## KEYWORDS

EST, Gallus gallus (chicken)

## SOURCE

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 822)  
AUTHORS Wittzell,H., Bed'Hom,B., Morin,V., Young,J.R., Whitaker,C.J.,  
Chausse,A.M. and Zoorob,R.  
TITLE A collection of chicken ESTs from activated immune cells  
JOURNAL Unpublished (2003)  
COMMENT Contact: Zoorob R  
UFR 1983  
CNRS

7 rue Guy Moquet, BP 8, 94801 Villejuif cedex, France  
Tel: 33 1 49 58 35 00  
Fax: 33 1 49 58 33 81  
Email: zoorob@jf.cnrs.fr.  
Location/Qualifiers

## FEATURES

## source

1..822  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9031"  
/cell\_type="Splenocytes"  
/clone\_id="PHA-activated splenocytes"  
/note="Vector: pTRIPLEX2"

## ORIGIN

## Alignment Scores:

Pred. No.:	5.75e-96	Length:	822
Score:	916.00	Matches:	175
Percent Similarity:	81.37%	Conservative:	39
Best Local Similarity:	66.54%	Mismatches:	49
Query Match:	66.14%	Indels:	1
DB:	7	Gaps:	0

US-10-649-273-2\_COPY\_148\_414 (1-267) x CF257246 (1-822)

QY 1 MetGluAlaHisAlaLeuThrLysArgLeuThrAsnLysValGlnPheProPheLeuVal 20  
DB 35 ATGGAGGCTCAGCAGCTTACATCACTGACAGACAGACAGCAATTTCTTCTTATGTT 94  
QY 21 LeuLeuLeuSerGlyLysCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 95 CTTTACTCTCGGAGGTACATGCATCTTGGCAGTAGACAGAGAGCTTTCAGATTTCCTT 154  
QY 41 LeuLeuGlyLysSerLeuAspPheAlaProGlyAspMetLeuAspLysValAlaArg 60  
DB 155 CTGCTTGACAGATCATAGATACACACAGGTGACATGTTGATTAAGTACAGAAAG 214  
QY 61 LeuSerLeuLeuLeuHisPheProLysCysSerThrMetSerGlyLysAlaLeuGlnHis 80  
DB 215 CTCTCTTAAAGAACACCCGAGGTCCACACATGCGCGGGGAAAGCAATAGAGCAC 274  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspPheLeuValProLeuHisAlaLys 100  
DB 275 CTGCTCAACCGGAGATCTGCGACAGATCACTTCAAGCTTCCATGCAACAGTATCGT 334  
QY 101 AsnCysAspPheSerPheThrgLysLeuGlnHisValThrAspLysLeuLeuMetLysLys 120  
DB 335 AACTGTGATTTTCTTTCTCGGACCTTGAGACCTTGCAACAAGCCATTCTTCAGAA 394  
QY 121 GlnLysGlnGlnGlyLysLeuLysGlnGlnLeuSerSerAlaAlaAspPheAlaAla 140  
DB 395 GAAAAAGAAAGGATTAACAAGAGGAAATCTGCTGCGTTAAGACATCTGCTGCT 454  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaLeuPhe 160  
DB 455 GCTGCACAGCAGTAGTGTGCTCTCATATTAATCAAGGAGACACAGGACCATGCTCTTC 514  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyLysVal 180  
DB 515 TGCATGAAAAACAGATATTAATACCAAAATCTGCAACTGTGTTATCAGAGAGAGT 574  
QY 181 AlaSerAsnPheThrgLysArgAlaLeuGlnLysLeuThrAsnAlaThrGlnCysThr 200  
DB 575 GCAAGTAATCATATATACGAAAGAGCTGACAGCTTGGCAAAATGCAACGTTTTCCT 634  
QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyLysLeuMetLeuAlaTrpAsnGly 220



done by E.M. Eddy, Ph.D. (National Institutes of Health, National Institute of Environmental Health Sciences).

Seq primer: 2069710

High quality sequence stop: 419.

#### FEATURES

location/Qualifiers

1..597

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="CD-1"

/db\_xref="taxon:10090"

/clone="IMAGE:6445750"

/sex="male"

/tissue\_type="spermatozoa, pooled from multiple mice"

/dev\_stage="60 day"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="McCarrey Eddy spermatozoa"

/note="Organ: testis; Vector: pBluescript SK-

(Stratagene); Site\_1: XhoI; Site\_2: EcoRI; cDNA oligo

dt-primed [5'-(GA)10-ACGAGTCGAGTTTCTTTT-3'] and

directionally cloned using 5' linkers 5'-AATTCGACGAC-3'

and 5'-CTCGCCG-3'. Size selection of >400bp material

gives average insert size ranging from 1-2 kb. Library was

mass excised (from lambda-unidirectional) and retransformed

single-stranded phagemids were prepped and transformed

into DH10B. Library contains 98% recombinants.

References: J. Androl. 20:635-639 and Gene 25:263-269.

Library constructed and donated by J. McCarrey, Ph.D.

(Southwest Foundation for Biomedical Research, Dept. of

Genetics); excision done by E.M. Eddy, Ph.D. (National

Institutes of Health, National Institute of Environmental

Health Sciences). Original lambda-based library is

available through ATCC, catalog #63422."

#### ORIGIN

##### Alignment Scores:

Pred. No.: 3,986-86 Length: 597  
Score: 830.00 Matches: 162  
Percent Similarity: 91.80% Conservative: 6  
Best Local Similarity: 88.52% Mismatches: 15  
Query Match: 59.93% Indels: 0  
DB: Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x CB272391 (1-597)

QY 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAlaValGlnPheProPheLeuVal 20  
DB 49 ATGAGGCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 108  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnValSerAspPheLeu 40  
DB 109 CTTTGAATTTCTGGCGGCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 168  
QY 41 LeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 60  
DB 169 CTCCTTGGAGAGTCTTGGAGATAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 228  
QY 61 LeuSerLeuIleLeuHisProGlnCysSerThrMetSerGlyGlyValAlaIleGlnHis 80  
DB 229 CTTTCTTAATCAAAACATCCAGAAATGTTCTCAATGAGTGGGAAAGCTATAGAACAG 288  
QY 81 LeuAlaIleGlnGlyAspArgPheHisPheAspIleLeuSerProPheLeuHisAlaIle 100  
DB 289 TTGGCCAAAGACGAAATAGATTCATTTTACTATCAATCCACTATGACAGATCTAG 348  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleIleMetLeuHis 120  
DB 349 AATTTGCAATTTTCTTTCACGGGACTTCACATATATCTGTAACCTAATTAACACACAG 408  
QY 121 GlnGlySerGlnGlyIleGlyLeuGlyGlnIleLeuSerSerAlaIleAspIleAlaIle 140  
DB 409 GAAAGACAGAGAGCATTGAGAGAGGGGCAATTCGTATCATGACGACATGCTCTCT 468

QY 141 ThrValGlnHisThrMetAlaCysHisLeuValIleArgArgThrHisArgAlaIleLeuPhe 160  
DB 469 GCGGTACAGACATGCAACAGCGTCCCTTGGCGAAAGAACACATCGCGCTATTCGTCTT 528  
QY 161 CysGlyGlnArgAspLeuLeuProGlnAlaGlnAlaValLeuValAlaSerGlyGlyVal 180  
DB 529 TCGACGACAGAAATTTGCTCTCTCCAGCTACGACGATTTGATGTGATGAGAGGTGT 588

QY 181 AlaSerAsn 183  
DB 589 GCAAGTAAAC 597

RESULT 13  
BU403563 792 bp mRNA linear EST 27-NOV-2002  
LOCUS BU403563  
DEFINITION 604138456F1 CSEQCHN59 Gallus gallus cDNA clone CHEST967014 5', mRNA  
sequence.

ACCESSION BU403563  
VERSION BU403563.1 GI:25772619  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 792)

BOARDMAN, P.E., SANZ-BEZUERO, J., OVERTON, I.M., BURT, D.W., BOSCH, E.,  
FONG, W.T., TICKLE, C., BROWN, W.R.A., WILSON, S.A. and HUBBARD, S.J.  
A Comprehensive Collection of Chicken cDNAs  
Curr. Biol. 12 (22), 1965-1969 (2002)

TITLE JOURNAL  
MEDLINE  
PUBMED  
COMMENT 12445392

Contact: Simon Hubbard  
Department of Biomedical Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.  
Location/Qualifiers

#### FEATURES

source

1..792

/organism="Gallus gallus"

/mol\_type="mRNA"

/strain="White Leghorn, Hixex"

/db\_xref="taxon:9031"

/clone="CHEST967014"

/dev\_stage="36"

/lab\_host="DH10B"

/clone\_lib="CSEQCHN59"

/note="Organ: limbs; Vector: pBluescript II KS(+); Site\_1:

EcoRI; Site\_2: NotI; This normalized library was

constructed from 1 million independent clones. cDNA

synthesis was initiated using an oligo(dT) primer, using

methyated C in the first strand synthesis reaction.

Following this first strand reaction, double-stranded cDNA

was blunt-ended, ligated to NotI adaptors, digested with

EcoRI, size-selected, and cloned into the NotI and EcoRI

compatible sites of a custom modified MCS of the

pBluescript (KS+) vector. The library was normalized in 2

rounds using conditions adapted from Soares et al., PNAS

(1994) 91: 9228-9232 and Bonaldi et al., Genome Research 6

(1996): 791, except that a significantly longer

reannealing hybridization was used."

#### ORIGIN

##### Alignment Scores:

Pred. No.: 1,016-84 Length: 792  
Score: 819.50 Matches: 164  
Percent Similarity: 78.99% Conservative: 39  
Best Local Similarity: 63.81% Mismatches: 53  
Query Match: 59.17% Indels: 3  
DB: Gaps: 1



## ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.

1 (bases 1 to 545)

## REFERENCE

Takauega, A., Hirotsune, S., Itoh, R., Jitchazon, A., Suzuki, H., Aso, H.  
and Sugimoto, Y.Establishment of a high throughput EST sequencing system using  
poly(A) tail-removed cDNA libraries and determination of 36,000  
bovine ESTs

## JOURNAL

MEDLINE

PUBMED

Nucleic Acids Res. 29 (22), E108 (2001)

## COMMENT

11713328

Contact: Yoshikazu Sugimoto

Animal Genetics Division

Shirakawa Institute of Animal Genetics

Odakura, Nihiyo, Nishi-shirakawa, Fukushima 961-8061, Japan

Tel: 81-248-25-5641

Fax: 81-248-25-5725

Email: kazueugi@cocoa.ocn.ne.jp

Single pass sequencing.

This clone was obtained from a polyA-deleted cDNA library.

Location/Qualifiers

## FEATURES

source

1..545

/organism="Bos taurus"

/mol\_type="mRNA"

/db\_xref="taxon:9913"

/clone="E1K1013A07"

/tissue\_type="kidney"

/dev\_stage="fetus"

/lab\_host="DH10B"

/clone\_id="Bos taurus kidney fetus"

/note="Vector: pZRI; Site\_1: SalI; Site\_2: NotI; Poly A  
was deleted from a NotI site"

## ORIGIN

## Alignment Scores:

Pred. No.:	2,14e-81	Length:	545
Score:	789.00	Matches:	147
Percent Similarity:	90.11%	Conservative:	17
Best Local Similarity:	80.77%	Mismatches:	17
Query Match:	56.97%	Indels:	1
DB:	1	Gaps:	0

US-10-649-273-2\_COPY\_148\_414 (1-267) x AV602901 (1-545)

QY	85	GLYAsnArgPheHisPheAspIleIleSerProIleuHisHis-AlaIysAsnCyAspPh	104
DB	544	GGAAATGATGATGCAATTTGATTCANCTCCATGCAACGGTCTTAATAATTTGATTT	485
QY	104	eserPheThrGlyIleuGlnHisValThrAspIleIleIleMetIysIysGluIysGluI	124
DB	484	TTCTTTTTCGACTTCACACGTTATGATGATGATGATGATGATGATGATGATGATGATG	425
QY	124	uGlyIleGlyIysGlyIleIleIleSerSerAlaAlaAspIleAlaIatThrValGlnHi	144
DB	424	AGGATATGACAGGAGGAGGAGCTCTGCTTCAGCTGCGGACATTCGCTTGGCGTTCAGCA	365
QY	144	srHMetAlaCyHisIleuValIysArgThrHisArgAlaIleIleuPheCyIysGlnAr	164
DB	364	CACGGTGGCTGCGACATTCGCAAAAAGAACACATGCTCTTCTGTTCTGCAAGCAGAG	305
QY	164	gaAspIleuProGlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPh	184
DB	304	AGCCTTCTTACATCAGAGTAACGACGATCTGTTGATCTGAGGCGCTGCAAGTAACCT	245
QY	184	eTyrlleArgArgAlaIleuGluIleIleuThrAsnAlaThrGlnCyThrIleuIleuCyAsPr	204
DB	244	ATATATCCGAAAGCCCTGGAAATTCGACCAATGCAACACAGTGCATTTGCTGTGCC	185
QY	204	oProProArgIleuCyThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLe	224
DB	184	GCCCCCAGACTCTGCACTGACCAACGCGGTATGATTCATGCAATGCTGTGAAGACT	125

QY	224	uArgAlaGlyIleuGlyIleIleuHisAspIleGlyIleArgTyrgluProIysCyAsPr	244
DB	124	ACGTGCTGGCTTGGGCAATTTACACAAACAGAGGACATCCGCTACGAACCAAAATGTC	65
QY	244	oleuGlyValaAspIleSerIysGluValGlyGluAlaSerIleIysValProGlnIleuIy	264
DB	64	TCTTGAGTATGATATATCAAAAAGAAAGTGGAGAGCTGCTATTAAGTGCACAGATTAA	5
QY	264	sMet 265	
DB	4	AAATG 1	

Search completed: November 11, 2005, 02:14:12  
Job time : 3793.43 secs

Mon Nov 14 07:47:35 2005

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:02:11 ; Search time 64.4261 Seconds  
(without alignments)  
1434.756 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 LIALVQGVSDFLKSLDI.....DISKEVGASIKVPOLKMEI 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1240	100.0	267	5	ABG96487 Novel hum
2	1240	100.0	414	5	ABG96478 Novel hum
3	1240	100.0	414	5	ABG96481 Human O-8
4	1240	100.0	414	6	ABJ26654 Human pro
5	1203	97.0	414	5	AAE29234 Human gly
6	1203	97.0	414	6	ABG71161 Novel hum
7	1203	97.0	414	6	ABU09569 Human gly
8	960	77.4	364	6	ADA54471 Human pro
9	332.5	26.8	409	4	ABG69133 Drosophil
10	326	26.3	463	3	AAV52216 Arabidops
11	319	25.7	245	3	AAV19287 Arabidops
12	319	25.7	443	3	AAV19286 Arabidops
13	319	25.7	444	3	AAV19285 Arabidops
14	257	20.7	382	6	ABU22934 Protein e
15	248.5	20.0	348	6	ABU35232 Protein e
16	247	19.9	312	6	ABU16575 Protein e
17	247	19.9	350	6	ADL05040 M. catarr
18	241	19.4	343	6	ABU39303 Protein e
19	239	19.3	251	5	AAE31054 Ehrlichia
20	231	18.6	341	4	AAU36205 Pseudomon
21	231	18.6	341	6	ABU38276 Protein e
22	231	18.6	341	7	ABU38276 Protein e
23	231	18.6	341	7	ABU38276 Protein e
24	229	18.5	341	7	ABU38276 Protein e
25	224	18.1	342	3	AAV52202 Haemophil

26	224	18.1	342	4	AAU35450	AAU35450 Haemophil
27	224	18.1	342	6	ABU30280	ABU30280 Protein e
28	213.5	17.2	347	4	ABE61112	ABE61112 Drosophil
29	213	17.2	337	4	AAU38187	AAU38187 Salimoneil
30	213	17.2	337	6	ABU47561	ABU47561 Protein e
31	212	17.1	340	6	ABU40514	ABU40514 Protein e
32	212	17.1	337	7	ADP06228	ADP06228 Bacteri
33	210	16.9	337	6	ABU50237	ABU50237 Protein e
34	209	16.9	326	4	AAE96423	AAE96423 Putative
35	208	16.8	335	6	ABU27480	ABU27480 Protein e
36	207.5	16.7	341	6	ABU40069	ABU40069 Protein e
37	206	16.6	337	3	AAV52204	AAV52204 Becherich
38	206	16.6	337	4	AAU34711	AAU34711 E. coli c
39	206	16.6	337	6	ABU28771	ABU28771 Protein e
40	205	16.5	343	7	ABO62704	ABO62704 Klebsiell
41	201.5	16.2	421	5	ABG96491	ABG96491 Novel hum
42	200	16.1	325	2	AAE26325	AAE26325 Glycoprot
43	200	16.1	325	3	AAV52203	AAV52203 Pasteurel
44	199.5	16.1	350	6	ABU55741	ABU55741 Protein e
45	199	16.0	354	6	ABU37844	ABU37844 Protein e

## ALIGNMENTS

RESULT 1	ABG96487	standard; protein; 267 AA.
ID	ABG96487	
AC	ABG96487	
DT	11-DEC-2002	(first entry)
DE	Novel human metalloprotease MPI fragment #1.	
KW	Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;	
KW	motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;	
KW	reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;	
KW	genital wart; metabolic disorder; premature puberty; Kallman syndrome;	
KW	Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;	
KW	Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;	
KW	liver disease; renal disease; immune disorder; rheumatoid arthritis;	
KW	acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;	
KW	emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;	
KW	neurological disorder.	
OS	Homo sapiens.	
PN	WO200272751-A2.	
PD	19-SEP-2002.	
PF	05-FEB-2002; 2002WO-US003353.	
PR	05-FEB-2001; 2001US-0266518P.	
PR	10-APR-2001; 2001US-0282814P.	
XX	(BRIM ) BRISTOL-MYERS SQUIBB CO.	
XX	Chen J, Feder J, Nelson TC, Duclos F, Krystek S;	
XX	WPI: 2002-723329/78.	
XX	N-PSDB; ABS76639.	
XX	New isolated nucleic acid encoding MP-1 protein, useful for preventing,	
XX	treating, or ameliorating diseases associated with aberrant	
XX	metalloproteinase activity, e.g. immune, metabolic, inflammatory and	
XX	neurological disorders.	
XX	Claim 5; Page 29; 473pp; English.	
XX	The invention describes an isolated nucleic acid molecule (I) encoding a	
XX	metalloprotease (MP-1). (I) is useful for preventing, treating, or	

CC ameliorating a medical condition, particularly an immune disorder, an  
 CC aberrant glutamate transport or motor neuron disorder, such as  
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
 CC condition. The compositions and methods are also useful for diagnosing,  
 CC prognosticating, treating, ameliorating and/or treating disorders  
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
 CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
 CC sequence of a metalloprotease MP1 protein

SQ Sequence 267 AA;

Query Match 100.0%; Score 1240; DB 5; Length 267;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-131;  
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIALVQGVSDPFLILGSLDIAPGDMLDKVARRLSLIKHPECSTWGGKAIIEHLAKQGNRF 60  
 DB 29 LIALVQGVSDPFLILGSLDIAPGDMLDKVARRLSLIKHPECSTWGGKAIIEHLAKQGNRF 88  
 QY 61 HFDIKRPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 120  
 DB 89 HFDIKRPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 148  
 QY 121 HLVKRTHRALILFCKQRDLIPONNAVIVASGVASNFYIRALIEILTNATQCTLLCPPRLL 180  
 DB 149 HLVKRTHRALILFCKQRDLIPONNAVIVASGVASNFYIRALIEILTNATQCTLLCPPRLL 208  
 QY 181 CTDNGIMIMWNGIERLRAGLGIILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239  
 DB 209 CTDNGIMIMWNGIERLRAGLGIILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 267

RESULT 2  
 ABG96478 ID ABG96478 standard; protein; 414 AA.  
 XX ABG96478;  
 AC  
 XX 11-DEC-2002 (first entry)  
 DT  
 XX  
 XX Novel human metalloprotease MP1.  
 DE  
 XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
 KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
 KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
 KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
 KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
 KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
 KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
 KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
 KW neurological disorder.  
 KW  
 XX Homo sapiens.  
 OS  
 XX  
 XX WO200272751-A2.  
 XX  
 XX 19-SEP-2002.  
 PD  
 XX  
 XX 05-FEB-2002; 2002MO-US003353.  
 PF  
 XX  
 XX 05-FEB-2001; 2001US-0266518P.  
 PR  
 XX 10-APR-2001; 2001US-0282814P.  
 XX  
 XX  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;  
 XX MPI; 2002-72329/78.  
 DR N-PSDB; ABS76635.  
 XX  
 PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
 PT treating, or ameliorating diseases associated with aberrant  
 PT metalloprotease activity, e.g. immune, metabolic, inflammatory and  
 PT neurological disorders.  
 XX  
 PS Claim 5; Fig 1A-C; 473pp; English.

CC The invention describes an isolated nucleic acid molecule (I) encoding a  
 CC metalloprotease (MP-1). (I) is useful for preventing, treating, or  
 CC ameliorating a medical condition, particularly an immune disorder, an  
 CC aberrant glutamate transport or motor neuron disorder, such as  
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
 CC condition. The compositions and methods are also useful for diagnosing,  
 CC prognosticating, treating, ameliorating and/or treating disorders  
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
 CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
 CC sequence of a metalloprotease MP1 protein

SQ Sequence 414 AA;

Query Match 100.0%; Score 1240; DB 5; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-130;  
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIALVQGVSDPFLILGSLDIAPGDMLDKVARRLSLIKHPECSTWGGKAIIEHLAKQGNRF 60  
 DB 176 LIALVQGVSDPFLILGSLDIAPGDMLDKVARRLSLIKHPECSTWGGKAIIEHLAKQGNRF 235  
 QY 61 HFDIKRPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 120  
 DB 236 HFDIKRPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 295  
 QY 121 HLVKRTHRALILFCKQRDLIPONNAVIVASGVASNFYIRALIEILTNATQCTLLCPPRLL 180  
 DB 296 HLVKRTHRALILFCKQRDLIPONNAVIVASGVASNFYIRALIEILTNATQCTLLCPPRLL 355  
 QY 181 CTDNGIMIMWNGIERLRAGLGIILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239  
 DB 356 CTDNGIMIMWNGIERLRAGLGIILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 414

RESULT 3  
 ABB05481 ID ABB05481 standard; protein; 414 AA.  
 AC ABB05481;  
 XX  
 XX 19-APR-2002 (first entry)  
 DT  
 XX  
 XX Human O-sialoglycoproteinase-like protein SEQ ID NO:2.  
 DE Human, O-sialoglycoproteinase-like protein; OSGPLP, enzyme.  
 KW  
 XX Homo sapiens.  
 OS  
 XX  
 XX CN1318550-A.  
 PN  
 XX  
 XX 24-OCT-2001.  
 PD  
 XX  
 XX 19-APR-2000; 2000CN-00106834.  
 PF  
 XX



PR 19-APR-2000; 2000CN-00106834.  
XX  
PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.  
PI Mao Y, Xie Y;  
XX  
DR WPI: 2002-115090/16.  
XX N-PSDB; ABA93268.  
PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful  
XX for diagnosing, preventing and treating related diseases.  
PS Claim 1; Page 30-31 (Disclosure); 38pp; Chinese.  
XX  
XX The present sequence represents human O-sialoglycoproteinase-like protein  
CC (OSGPLP). The present invention also describes: (1) the preparation of  
CC the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the  
CC prevention and/or treatment of related diseases; (4) utilizing the OSGPLP  
CC protein in screening its agonist, excitomotor and inhibitor and preparing  
CC an antibody against the OSGPLP protein; and (5) the use of the OSGPLP  
CC polynucleotide sequences, proteins, agonists, excitomotors, inhibitors  
CC and antibodies in treating diseases related to the abnormal OSGPLP gene  
CC and in preparing the medicine composite for the treatment  
XX  
SQ Sequence 414 AA;  
Query Match 100.0%; Score 1240; DB 5; Length 414;  
Best Local Similarity 100.0%; Pred. No. 1.5e-130;  
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LIALVGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGKAIETHLAKGNRF 60  
DB 176 LIALVGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGKAIETHLAKGNRF 235  
QY 61 HFDIKPPLHAKNCDFSTGLOHTVDKTIIMKEKEEGIEKGQILSSAADIAATVQHTMAC 120  
DB 236 HFDIKPPLHAKNCDFSTGLOHTVDKTIIMKEKEEGIEKGQILSSAADIAATVQHTMAC 295  
QY 121 HLVKRTTRAILFCQKORDLLPQNNNAVLVASGVASNFYIRRALEILTNAOTCTLLCPPRL 180  
DB 296 HLVKRTTRAILFCQKORDLLPQNNNAVLVASGVASNFYIRRALEILTNAOTCTLLCPPRL 355  
QY 181 CTDNGIMIANWGIERLRAGILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239  
DB 356 CTDNGIMIANWGIERLRAGILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 414  
RESULT 4  
ABJ26654  
ID ABJ26654 standard; protein; 414 AA.  
XX  
AC ABJ26654;  
XX  
DT 01-MAY-2003 (first entry)  
XX  
DE Human protein modification + maintenance molecule protein SEQ ID No 8.  
XX  
XX Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;  
KM cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;  
KM antitumor; hepatotropic; gynaecological; antibacterial; virucide;  
KM protooncogene; antiparasitic; cell proliferative disease; PMOD;  
KM protein modification and maintenance molecule; immunogenic fragment;  
KM cancer; autoimmune; inflammatory disease; neurological disorder;  
KM gastrointestinal; developmental; vesicle trafficking disorder; infection;  
KM protein-protein interaction; drug-target interaction;  
KM gene expression profile; human.  
XX  
OS Homo sapiens.  
XX  
PN WC2003000844-A2.  
XX  
PD 03-JAN-2003.

PF 18-JUN-2002; 2002WO-US019360.  
XX  
XX 22-JUN-2001; 2001US-0300508P.  
PR 06-JUL-2001; 2001US-0303445P.  
PR 13-JUL-2001; 2001US-0305405P.  
PR 09-AUG-2001; 2001US-0311442P.  
PR 24-AUG-2001; 2001US-0314621P.  
PR 29-AUG-2001; 2001US-0315992P.  
PR 03-MAY-2002; 2002US-0378205P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
XX Gandhi AR, Kable AE, Swarnakar A, Hafalia ALA, Tran B, Duggan BM,  
PI Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DAM, Lee EA, Yue H;  
PI Foraythe J, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J;  
PI Thangavelu K, Gietzen KJ, Ding L, Baughn MR, Borowsky ML, Yao MG;  
PI Walla NK, Mason PM, Gunturjan R, Lee S, Becha SD, Lee SY, Tran UK;  
PI Elliott VS, Luo W, Sprague MW, Tang YT, Lu Y, Zebardadian Y;  
XX  
DR WPI: 2003-184039/18.  
XX N-PSDB; ABR23207.  
XX  
PT New isolated human PMOD polypeptide and polynucleotide, useful for  
PT diagnosing, treating and preventing diseases or conditions associated  
PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and  
PT infections.  
XX  
PS Claim 63; Page 182-183; 225pp; English.  
XX  
XX The invention relates to an isolated polypeptide comprising: any of 28  
CC sequences of 48-1256 amino acids; a natural amino acid sequence at least  
CC 90% identical to the 28 amino acid sequences, 94% identical to a sequence  
CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino  
CC acids, or 97% identical to a sequence of 242 amino acids, all given in  
CC the specification; or a biologically active or immunogenic fragment of  
CC the isolated polypeptide. The polypeptides and polynucleotides are useful  
CC in diagnosing, treating and preventing diseases or conditions associated  
CC with the decreased expression of protein modification and maintenance  
CC molecules (PMOD), such as cell proliferative diseases (e.g. cancer,  
CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,  
CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,  
CC epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.  
CC endometriosis), developmental, vesicle trafficking disorders, and  
CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also  
CC useful in assessing the effects of exogenous compounds on the expression  
CC of nucleic acid and amino acid sequences of PMOD. The PMOD or its  
CC fragments are useful in screening compounds for effectiveness as agonist  
CC or antagonist of the polypeptides, or in altering the expression of the  
CC target polynucleotide and compounds that specifically bind to or modulate  
CC the activity of the polypeptide. The microarray is useful in monitoring  
CC or measuring protein-protein interactions, drug-target interactions, and  
CC gene expression profiles. This sequence represents a human PMOD protein  
CC of the invention  
XX  
SQ Sequence 414 AA;  
Query Match 100.0%; Score 1240; DB 6; Length 414;  
Best Local Similarity 100.0%; Pred. No. 1.5e-130;  
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LIALVGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGKAIETHLAKGNRF 60  
DB 176 LIALVGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMSGKAIETHLAKGNRF 235  
QY 61 HFDIKPPLHAKNCDFSTGLOHTVDKTIIMKEKEEGIEKGQILSSAADIAATVQHTMAC 120  
DB 236 HFDIKPPLHAKNCDFSTGLOHTVDKTIIMKEKEEGIEKGQILSSAADIAATVQHTMAC 295  
QY 121 HLVKRTTRAILFCQKORDLLPQNNNAVLVASGVASNFYIRRALEILTNAOTCTLLCPPRL 180  
DB 296 HLVKRTTRAILFCQKORDLLPQNNNAVLVASGVASNFYIRRALEILTNAOTCTLLCPPRL 355  
QY 181 CTDNGIMIANWGIERLRAGILHDIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239

Db	356	CTDNGIMTAMNGIERLAGLILHDIEGIRYERKCPGLVDISKVEGSAISIKPOLKMEI	414
RESULT 5			
ID	AAE29234	AAE29234 standard; protein; 414 AA.	
AC	AAE29234;		
XX			
DT	27-JAN-2003	(first entry)	
XX			
DE	Human glycoprotease 28472 protein.		
XX			
KW	Human; adenosine deaminase; seven transmembrane domain receptor; cancer;		
KW	7TM; glycoprotease; immune disorder; IGA deficiency; allergy; arhythmia		
KW	rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder		
KW	hypertension; ischaemic heart disease; obesity; myocardial infarction;		
KW	endothelial cell disorder; Grave's disease; psoriasis; brain disorder;		
KW	Parkinson's disease; Alzheimer's disease; haematopoietic disorder;		
KW	cerebral oedema; metabolic disorder; liver disorder; platelet disorder;		
KW	chromosome mapping; tissue typing; gene therapy; neuroprotective;		
KW	cytostatic; anorectic; cardiac; haemostatic.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	Domain	1..108	
FT		/note="Non-transmembrane domain; N-terminal cytoplasmic	
FT		domain"	
FT	Domain	38..369	
FT		/note="Endopeptidase O-sialoglycoprotein hydrolase	
FT		metalloprotease zinc glycoprotease sialoglycoprotease	
FT		domain"	
FT	Domain	109..132	
FT		/note="Transmembrane domain"	
FT	Domain	133..164	
FT		/note="Non-transmembrane domain; non-cytoplasmic loop"	
FT	Domain	138..152	
FT		/note="Glycoprotease domain"	
FT	Domain	165..189	
FT		/note="Transmembrane domain"	
FT	Domain	190..316	
FT		/note="Non-transmembrane domain; cytoplasmic domain"	
FT	Domain	317..333	
FT		/note="Transmembrane domain"	
FT	Domain	334..414	
FT		/note="Non-transmembrane domain"	
FT	Domain	374..414	
FT		/note="Sialoglycoprotease type domain"	
XX			
PN	WO200274960-A2.		
PD	26-SEP-2002.		
XX			
PF	08-NOV-2001; 2001WO-US051427.		
XX			
PR	08-NOV-2000; 2000US-0246768P.		
PR	08-NOV-2000; 2000US-0246772P.		
PR	15-NOV-2000; 2000US-0249185P.		
XX			
PA	(MILL-) MILLENNIUM PHARM INC.		
XX			
PI	Leibyl KR, Kapeller-Libermann R, Glucksmann M;		
XX			
DR	WPI; 2002-759898/82.		
DR	N-PSDB; AAD46856.		
XX			
PT	New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,		
PT	useful for diagnosing and treating cancer, immune, cardiovascular,		
PT	hematopoietic, brain, pain, metabolic, liver or platelet disorders, and		
PT	in pharmacogenomics.		
XX			

XX	PS	Claim 1, Fig 8; 178pp; English.
XX	CC	The present invention relates to novel 38650, 28472, 5495, 65507, 81588
XX	CC	or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-
XX	CC	protease or seven transmembrane domain (7TM) receptor family members.
XX	CC	Sequences of the invention are useful in diagnosing and treating cancer
XX	CC	or aberrant cellular proliferation and/or differentiation (e.g. colon or
XX	CC	lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid
XX	CC	arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,
XX	CC	hypertension, atherosclerosis, arrhythmias, ischemic heart disease,
XX	CC	myocardial infarction, thrombus) including endothelial cell disorders
XX	CC	(e.g. psoriasis, Grave's disease), hematopoietic disorders, brain
XX	CC	disorders (e.g. cerebral edema, Parkinson's or Alzheimer's disease),
XX	CC	pain and metabolic disorders (e.g. obesity), liver disorders or platelet
XX	CC	disorders. They are also useful in screening assays, predictive medicine
XX	CC	(e.g. diagnostic assays, prognostic assays, monitoring clinical trials
XX	CC	and pharmacogenetics) and prophylactic and therapeutic methods. The
XX	CC	nucleic acids may also be used in chromosome mapping, tissue typing and
XX	CC	forensic biology and as surrogate markers. Sequences of the invention are
XX	CC	also used in gene therapy. The present sequence is human glycoprotease
XX	CC	28472 protein
SQ		Sequence 414 AA;
Query Match	97.0%;	Score 1203; DB 5; Length 414;
Best Local Similarity	97.1%;	Pred. No. 2,36-126;
Conservative	3;	Mismatches 4; Indels 0; Gaps 0;
QY	1	LLALVQGVSDLLGKSLIDTAPGDMLKVARRLSLIKHPECTSWGKATIEHLAKGNRF 60
DB	176	LLALVQGVSDLLGKSLIDTAPGDMLKVARRLSLIKHPECTSWGKATIEHLAKGNRF 235
QY	61	HPDIKPLPHAKNCDPFTGLOHTYTDKIMKEKEBEGIEKQIILSSADIAATVQHTMAC 120
DB	236	HPDIKPLPHAKNCDPFTGLOHTYTDKIMKEKEBEGIEKQIILSSADIAATVQHTMAC 295
QY	121	HLVYKTRTHALIFCKQRIILPQNNALVYASGVANFPIYIRALLETITNAOTCTLLCPPRL 180
DB	296	HLVYKTRTHALIFCKQRIILPQNNALVYASGVANFPIYIRALLETITNAOTCTLLCPPRL 355
QY	181	CTDNGIMANNGIERLRAGILHIDIEIRYEPKCPGVDSIKVEGASIKVPOLKMEI 239
DB	356	CTDNGIMANNGIERLRAGILHIDIEIRYEPKCPGVDSIKVEGASIKVPOLKMEI 414
RESULT 6		
ABG71161		
ID	ABG71161	standard; protein; 414 AA.
XX	AC	ABG71161;
XX	AC	
XX	DT	30-JAN-2003 (first entry)
XX	DE	Novel human glycoprotease 28472.
XX	KW	Cancer; aberrant cell proliferation; aberrant cell differentiation;
XX	KW	breast cancer; ovarian cancer; prostate cancer; colon cancer;
XX	KW	lung cancer; immune disorder; heart disorder; cardiovascular disorder;
XX	KW	endothelial disorder; hematopoietic disorder; blood vessel disorder;
XX	KW	brain disorder; pain; metabolic disorder; liver disorder; diabetes;
XX	KW	platelet disorder; carcinoma; sarcoma; leukaemia; Hodgkin's disease;
XX	KW	autoimmune disorder; hypertension; atherosclerosis; heart failure;
XX	KW	myocardial infarction; ischemic heart disease; Crohn's disease;
XX	KW	Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;
XX	KW	cerebral ischaemia; peripheral neuropathy; Alzheimer's disease;
XX	KW	Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; enzyme.
OS		Homo sapiens.
XX	FT	Key
XX	FT	Location/Qualifiers
XX	FT	138..152
XX	FT	/label=Glycoprotease_domain

PN WO00277233-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 08-NOV-2001; 2001WO-US046724.  
XX  
PR 08-NOV-2000; 2000US-0246768P.  
PR 08-NOV-2000; 2000US-0246772P.  
PR 15-NOV-2000; 2000US-0249185P.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
PI Leiby KR, Kapeller-Libermann R, Glucksmann M;  
XX WPI; 2003-029938/02.  
DR N-PSDB; ABB57020.  
XX  
PT New adenosine deaminase, glycoprotease and seven transmembrane domain  
PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,  
PT 81588 and 14354, useful for treating e.g. leukemia, Hodgkin's disease or  
PT hypertension.  
XX  
PS Claim 4; Fig 8A-B; 178pp; English.  
XX  
XX The invention describes isolated 38650, 28472, 5495, 65507, 81588 and  
CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The  
CC 38650 nucleic acid molecule comprises a sequence encoding adenosine  
CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding  
CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise  
CC sequences that encode a human seven transmembrane domain (7TM). The  
CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide  
CC sequences are useful for diagnosing, preventing or treating a subject  
CC with or at risk of developing a disorder, e.g. cancer or aberrant  
CC cellular proliferation and/or differentiation (e.g. breast, ovarian,  
CC prostate, colon or lung cancer), immune disorders, heart disorders,  
CC cardiovascular disorders, endometrial disorders, hematopoietic disorders,  
CC blood vessel disorders, brain disorders, pain and metabolic disorders,  
CC liver disorders or platelet disorders. These disorders include carcinoma,  
CC sarcoma, leukaemia, Hodgkin's disease, autoimmune disorders,  
CC hypertension, atherosclerosis, heart failure, myocardial infarction,  
CC ischemic heart disease, Crohn's disease, Grave's disease, Kawasaki  
CC syndrome, Raynaud's disease, aneurysm, cerebral ischaemia, peripheral  
CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,  
CC cecelia or diabetes. This is the amino acid sequence of the novel human  
CC glycoprotease 28472  
XX  
SQ Sequence 414 AA;  
Query Match 97.0%; Score 1203; DB 6; Length 414;  
Best Local Similarity 97.1%; Pred. No. 2,3e-126;  
Matches 232; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 1 LLAIVGVSDPFLKSKSLDIAPGMDLVKVARRLSLIKHPCSTMSGKRAIEHLAKQGNRF 60  
DB 176 LLAIVGVSDPFLKSKSLDIAPGMDLVKVARRLSLIKHPCSTMSGKRAIEHLAKQGNRF 235  
QY 61 HFDIKRPLHAKNCDFSTFGQHTVDKTIIMKKEKEGIEKGQIISASADIAATVQHTMAC 120  
DB 236 HFDIKRPLHAKNCDFSTFGQHTVDKTIIMKKEKEGIEKGQIISASADIAATVQHTMAC 295  
QY 121 HLVRTRRAILFCQKORDLLPNNNAVLVAGGVASFYRRLALEITLTNAQCTLCPPL 180  
DB 296 HLVRTRRAILFCQKORDLLPNNNAVLVAGGVASFYRRLALEITLTNAQCTLCPPL 355  
QY 181 CTDNGMIAMNGIERLRAGLILHDIEGIRYRKPCLGVDSKEVGEASIKVPOLKMEI 239  
DB 356 CTDNGMIAMNGIERLRAGLILHDIEGIRYRKPCLGVDSKEVGEASIKVPOLKMEI 414

AC AEU09569;  
XX  
DT 08-JUL-2003 (first entry)  
XX  
DE Human glycoprotease encoded by cDNA 28472.  
XX  
XX Human, enzyme; cancer; aberrant cellular proliferation; differentiation;  
KW immune disorders; heart disorder; brain disorder;  
KW cardiovascular disorder; endometrial cell disorder; pain disorder;  
KW hematopoietic disorder; blood vessel disorder; metabolic disorder;  
KW liver disorder; platelet disorder; glycoprotease.  
XX  
OS Homo sapiens.  
XX  
PN US2003009017-A1.  
XX  
XX  
PD 09-JUN-2003.  
XX  
PF 08-NOV-2001; 2001US-00012140.  
XX  
PR 08-NOV-2000; 2000US-0246768P.  
PR 08-NOV-2000; 2000US-0246772P.  
PR 15-NOV-2000; 2000US-0249185P.  
XX  
XX (LEIB/) LEIBY K R.  
PA (KAP/) KAPPELLER-LIBERMANN R.  
PA (GLUC/) GLUCKSMANN M A.  
XX  
PI Leiby KR, Kapeller-Libermann R, Glucksmann MA;  
XX WPI; 2003-428888/40.  
DR N-PSDB; ACA60887.  
XX  
PT New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid  
PT molecules, useful for diagnosing, treating cancer, pain, or immune,  
PT heart, endometrial cell, hematopoietic, blood vessel, brain, metabolic  
PT and liver disorders.  
XX  
PS Claim 4; Fig 8; 90pp; English.  
XX  
XX The invention relates to an isolated 38650 (encoding adenosine  
CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7  
CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or  
CC a sequence which is at least 60% identical to the six nucleic acids or  
CC their open reading frames, fragments of at least 15 nucleotides,  
CC naturally occurring variants, or a DNA insert of the plasmid deposited  
CC with the American Type Culture Collection as Accession No. not defined in  
CC the specification, which encodes the amino acid sequence). Also included  
CC are a host cell containing the nucleic acids (used to produce the  
CC protein), the encoded protein, an antibody that selectively binds to  
CC the polypeptide, and identifying a compound that binds to/modulates the  
CC activity of the polypeptide. The nucleic acid molecules, polypeptides and  
CC methods are useful for diagnosing, treating cancer, aberrant cellular  
CC proliferation and/or differentiation, immune disorders, heart disorders,  
CC cardiovascular disorders including endometrial cell disorders,  
CC haematopoietic disorders, blood vessel disorders, brain disorders, pain  
CC and metabolic disorders, liver disorders and platelet disorders (many  
CC examples of these disorders are given in the specification). The present  
CC sequence is the Human glycoprotease encoded by cDNA 28472  
XX  
SQ Sequence 414 AA;  
Query Match 97.0%; Score 1203; DB 6; Length 414;  
Best Local Similarity 97.1%; Pred. No. 2,3e-126;  
Matches 232; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 1 LLAIVGVSDPFLKSKSLDIAPGMDLVKVARRLSLIKHPCSTMSGKRAIEHLAKQGNRF 60  
DB 176 LLAIVGVSDPFLKSKSLDIAPGMDLVKVARRLSLIKHPCSTMSGKRAIEHLAKQGNRF 235  
QY 61 HFDIKRPLHAKNCDFSTFGQHTVDKTIIMKKEKEGIEKGQIISASADIAATVQHTMAC 120  
DB 236 HFDIKRPLHAKNCDFSTFGQHTVDKTIIMKKEKEGIEKGQIISASADIAATVQHTMAC 295

Qy 121 HLVKRTHRAILFCQKORDLLPQNNAVLVASGGVSNFYIRRALEILTNAQTCTLLCPPRL 180  
 |||||  
 Db 296 HLVKRTHRAILFCQKORDLLPQNNAVLVASGGVSNFYIRRALEILTNAQTCTLLCPPRL 355  
 |||||  
 Qy 181 CTDNGIMTANNGIERLRAGLGLHDIBGIRYEPKCPDGLVDISKVEGASIKVPOLKMEI 239  
 |||||  
 Db 356 CTDNGIMTANNGIERLRAGLGLHDIBGIRYEPKCPDGLVDISKVEGASIKVPOLKMEI 414  
 |||||

## RESULT 8

ADAS4471  
 ID ADA54471 standard; protein; 364 AA.  
 XX  
 AC ADA54471;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Human protein, SEQ ID 2039.  
 XX  
 KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Noctropic;  
 KM Gene therapy; human; secretory protein; membrane proteins; cancer;  
 KM Inflammatory disease; osteoporosis; neurological disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI293569-A2.  
 XX  
 PD 19-MAR-2003.  
 XX  
 PF 21-MAR-2002; 2002EP-00006586.  
 XX  
 PR 14-SEP-2001; 2001JP-00328381.  
 PR 24-JAN-2002; 2002US-0350435P.  
 XX  
 PA (HELI-) HELIX RES INST.  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuno Y;  
 XX  
 DR WPI; 2003-395539/38.  
 DR N-PSDB; ADA52832.  
 XX  
 PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
 PT and/or membrane proteins, useful for developing medicines for diseases in  
 PT which the gene is involved, or as target molecules for gene therapy.  
 XX  
 PS Claim 14; SEQ ID NO 2039; 205pp; English.  
 XX  
 CC The present invention relates to novel human secretory or membrane  
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
 CC ADA54071). The coding sequences are useful in the gene therapy of  
 CC diseases caused by abnormalities of the proteins, e.g. cancer,  
 CC inflammatory diseases, osteoporosis or neurological disease.  
 XX  
 SQ Sequence 364 AA;

Query Match 77.4%; Score 960; DB 6; Length 364;  
 Best Local Similarity 98.4%; Pred. No. 5e-99;  
 Matches 186; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LLAIVGVSDPFLGKSLDIAPGDMIDKVARRLSLIKHPECSMTSGGKAI EHLAKQGNRF 60  
 |||||  
 Db 176 LLAIVGVSDPFLGKSLDIAPGDMIDKVARRLSLIKHPECSMTSGGKAI EHLAKQGNRF 235  
 |||||  
 Qy 61 HPDIKPLHNAKNCPSFTGLQHTVDKIMKKEKEGIEKGQILSSAADIATVQHTMAC 120  
 |||||  
 Db 236 HPDIKPLHNAKNCPSFTGLQHTVDKIMKKEKEGIEKGQILSSAADIATVQHTMAC 295  
 |||||  
 Qy 121 HLVKRTHRAILFCQKORDLLPQNNAVLVASGGVSNFYIRRALEILTNAQTCTLLCPPRL 180  
 |||||

Db 296 HLVKRTHRAILFCQKORDLLPQNNAVLVASGGVSNFYIRRALEILTNAQTCTLLCPPRL 355  
 |||||  
 Qy 181 CTDNGIMTAN 189  
 |||||  
 Db 356 CTDNGIMTAN 364  
 |||||

## RESULT 9

ABB69133  
 ID ABB69133 standard; protein; 409 AA.  
 XX  
 AC ABB69133;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 34191.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PMD, Myers EW;  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL13236.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 34191; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 409 AA;

Query Match 26.8%; Score 332.5; DB 4; Length 409;  
 Best Local Similarity 35.9%; Pred. No. 3.9e-28;  
 Matches 83; Conservative 38; Mismatches 97; Indels 13; Gaps 5;

Qy 2 LLAIVGVSDPFLGKSLDIAPGDMIDKVARRLSLIKHPECSMTSGGKAI EHLAK-QGNRF 60  
 |||||  
 Db 167 LVAANPGRVTLTGQTLDDPGAPFKIGRLHLILPYRLMNGRAI EHLAAQLASDPL 226  
 |||||  
 Qy 61 HPDIKPLHNAKNCPSFTGLQHTVDKIMKKEKEGIEKGQILSSAADIATVQHTMAC 120  
 |||||  
 Db 227 AYRFPPLAQQNCNSFPAGIKNSFRATRAARARATPTPDGYISNVDGCAGLRSVSR 286  
 |||||  
 Qy 121 HLVKRTHRAILFCQKORDLLPQNNAVLVASGGVSNFYIRRALEILTNAQTCTLLCPPRL 178  
 |||||  
 Db 287 HLKHFRTORALIEVCLLPHRQLFGDTPPTLVMSGGVANNDALYANI EHLAAQYGRSFRPSK 346  
 |||||

QY 179 RLCTDNGIMIAMNGIERLRAGILHDIE-GIRYEPKCPGLVDISKVEGEA 228  
DB 347 RYCSDNCGVMWMTAMHGVEQL-----LQDKEASTRYDYD--SIDIQSAGFA 388

RESULT 10  
AAVS2216  
ID AAVS2216 standard; protein; 463 AA.  
XX  
XX AAVS2216;  
XX  
XX 09-FEB-2000 (first entry)  
XX  
XX Arabidopsis thaliana ysjd protein homologue.  
XX  
XX ysjd protein; essential; Gram positive; Gram negative; conserved; motif;  
XX identification; antagonist; antibacterial; antibiotic; broad spectrum;  
XX treatment; infection; resistance; drug target.  
XX  
XX Arabidopsis thaliana.  
XX  
XX  
XX Key Location/Qualifiers  
XX FH 86..96  
XX FT /note="ysjd conserved motif 3"  
XX FT 111..131  
XX FT Region /note="ysjd conserved motif 4"  
XX FT 152..198  
XX FT Region /note="ysjd conserved motif 2"  
XX FT 208..259  
XX FT Region /note="ysjd conserved motif 1"  
XX  
XX PN W09954470-A2.  
XX  
XX PD 28-OCT-1999.  
XX  
XX PF 20-APR-1999; 99WO-EP002635.  
XX  
XX PR 22-APR-1998; 98GB-00008423.  
XX  
XX PA (GLAXO ) GLAXO GROUP LTD.  
XX  
XX PI Arigoni F, Edgerton MD, Loferer H, Peitsch MC;  
XX  
XX DR WPI; 2000-013253/01.  
XX  
XX PT Novel bacterial polypeptides used to identify broad spectrum antibiotics.  
XX  
XX PS Claim 1; Fig 1; 55pp; English.  
XX  
XX Proteins AAVS2202-Y52217 encompasses a novel family of proteins designated  
XX the ysjd family, after the name given to the Escherichia coli family  
XX member. These proteins are essential for the survival of both Gram  
XX negative and Gram positive bacteria, although no function has as yet been  
XX ascribed to these proteins. The ysjd proteins, fragments of ysjd proteins  
XX (for example, fragments encompassing one or more conserved ysjd motifs  
XX such as AAVS2218-Y52284) and nucleotides encoding them can be used to  
XX identify antagonists and broad spectrum antibacterial compounds. These  
XX antagonists and compounds can be used to treat a wide range of bacterial  
XX infections. New antibiotics are urgently needed, as serious bacterial  
XX infections and antibiotic resistant strains are becoming increasingly  
XX prevalent. The proteins of the invention are essential proteins for  
XX bacterial viability, and represent new targets for antibiotics

SQ Sequence 463 AA;

Query Match 26.3%; Score 326; DB 3; Length 463;  
Best Local Similarity 33.6%; Pred. No. 2.6e-27;  
Matches 82; Conservative 34; Mismatches 84; Indels 44; Gaps 5;

QY 1 LLAIVGCVSPDLLIGKCIDIAFGDNDKRVARRSLGIIINFGFCSTISGCAIEHIAAQQKRF 60  
DB 221 LLAIVAHKLGQYTOGTTVDVAIGBAFDXTAKMGLDWH-----RSGGPAVELALLEGDAK 275

QY 61 HEDIKRPLHAACNCPSPFGLQHWVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 120  
DB 276 SVKFNVPMPKXHDQCNFSYAGLKTQYRLAIEAKE-----IRNRADIASFQVAVL 325

QY 121 HLKRTTRAILPCKQBDLLPONNAVIVASGVASNFYIRRALEILTNACQCTLCPPPR 180  
DB 326 HLEEKCEKRAIDWALE--LEPSIKHWIVISGVASNKYVRLRLNIVENKYLKLVCPPSL 382

QY 181 CTDNGIMIAMNGIERLRAGILHDIEGIRYE-----PKCPGLVDISKE 224  
DB 383 CTDNGVMWMTAMHGVEHRRVQ-----RYDPPPTATPEDDYVYDLRRPRWPLGEERYAKG 432

QY 225 VGEA 228  
DB 433 RSEA 436

RESULT 11  
AAG19287  
ID AAG19287 standard; protein; 245 AA.  
XX  
XX AAG19287;  
XX  
XX DT 17-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 21031.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
XX Arabidopsis thaliana.  
XX  
XX PN EP1033405-A2.  
XX  
XX PD 06-SEP-2000.  
XX  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX  
XX PR 25-FEB-1999; 99US-0121825P.  
XX PR 05-MAR-1999; 99US-0123180P.  
XX PR 09-MAR-1999; 99US-0123548P.  
XX PR 23-MAR-1999; 99US-0125788P.  
XX PR 25-MAR-1999; 99US-0126264P.  
XX PR 29-MAR-1999; 99US-0126785P.  
XX PR 01-APR-1999; 99US-0127462P.  
XX PR 06-APR-1999; 99US-0128234P.  
XX PR 08-APR-1999; 99US-0128714P.  
XX PR 16-APR-1999; 99US-0129845P.  
XX PR 19-APR-1999; 99US-0130077P.  
XX PR 21-APR-1999; 99US-0130449P.  
XX PR 23-APR-1999; 99US-0130510P.  
XX PR 28-APR-1999; 99US-0130891P.  
XX PR 30-APR-1999; 99US-0131449P.  
XX PR 30-APR-1999; 99US-0132048P.  
XX PR 04-MAY-1999; 99US-0132407P.  
XX PR 05-MAY-1999; 99US-0132484P.  
XX PR 06-MAY-1999; 99US-0132485P.  
XX PR 06-MAY-1999; 99US-0132486P.  
XX PR 07-MAY-1999; 99US-0132487P.  
XX PR 11-MAY-1999; 99US-0132863P.  
XX PR 11-MAY-1999; 99US-0134256P.  
XX PR 14-MAY-1999; 99US-0134218P.  
XX PR 14-MAY-1999; 99US-0134219P.  
XX PR 14-MAY-1999; 99US-0134221P.  
XX PR 14-MAY-1999; 99US-0134370P.  
XX PR 18-MAY-1999; 99US-0134768P.  
XX PR 19-MAY-1999; 99US-0134941P.  
XX PR 20-MAY-1999; 99US-0135124P.  
XX PR 21-MAY-1999; 99US-0135353P.  
XX PR 24-MAY-1999; 99US-0135629P.  
XX PR 25-MAY-1999; 99US-0136021P.  
XX PR 27-MAY-1999; 99US-0136392P.

PR	2-MAY-1999	99US-01367882Z
PR	01-JUN-1999	99US-0137228Z
PR	03-JUN-1999	99US-0137528Z
PR	04-JUN-1999	99US-0137502Z
PR	07-JUN-1999	99US-0137724Z
PR	08-JUN-1999	99US-0138094Z
PR	10-JUN-1999	99US-0138540Z
PR	10-JUN-1999	99US-0138847Z
PR	14-JUN-1999	99US-0139119Z
PR	16-JUN-1999	99US-0139452Z
PR	17-JUN-1999	99US-0139453Z
PR	18-JUN-1999	99US-0139492Z
PR	18-JUN-1999	99US-0139454Z
PR	18-JUN-1999	99US-0139455Z
PR	18-JUN-1999	99US-0139456Z
PR	18-JUN-1999	99US-0139457Z
PR	18-JUN-1999	99US-0139458Z
PR	18-JUN-1999	99US-0139459Z
PR	18-JUN-1999	99US-0139646Z
PR	18-JUN-1999	99US-0139641Z
PR	18-JUN-1999	99US-0139642Z
PR	18-JUN-1999	99US-0139643Z
PR	22-JUN-1999	99US-0139698Z
PR	22-JUN-1999	99US-0140353Z
PR	22-JUN-1999	99US-0140354Z
PR	24-JUN-1999	99US-0140695Z
PR	28-JUN-1999	99US-0140982Z
PR	28-JUN-1999	99US-0140981Z
PR	30-JUN-1999	99US-0141287Z
PR	01-JUL-1999	99US-0141842Z
PR	01-JUL-1999	99US-0142154Z
PR	01-JUL-1999	99US-0142055Z
PR	06-JUL-1999	99US-0142290Z
PR	06-JUL-1999	99US-0142803Z
PR	09-JUL-1999	99US-0142920Z
PR	12-JUL-1999	99US-0142977Z
PR	15-JUL-1999	99US-0143542Z
PR	15-JUL-1999	99US-0144005Z
PR	16-JUL-1999	99US-0144086Z
PR	19-JUL-1999	99US-0144325Z
PR	19-JUL-1999	99US-0144331Z
PR	19-JUL-1999	99US-0144332Z
PR	19-JUL-1999	99US-0144333Z
PR	19-JUL-1999	99US-0144334Z
PR	21-JUL-1999	99US-0144508Z
PR	21-JUL-1999	99US-0145088Z
PR	22-JUL-1999	99US-0145087Z
PR	22-JUL-1999	99US-0145089Z
PR	22-JUL-1999	99US-0145192Z
PR	23-JUL-1999	99US-0145195Z
PR	23-JUL-1999	99US-0145218Z
PR	23-JUL-1999	99US-0145224Z
PR	26-JUL-1999	99US-0145376Z
PR	27-JUL-1999	99US-0145513Z
PR	27-JUL-1999	99US-0145518Z
PR	28-JUL-1999	99US-0145519Z
PR	28-JUL-1999	99US-0145591Z
PR	02-AUG-1999	99US-0146386Z
PR	02-AUG-1999	99US-0146388Z
PR	02-AUG-1999	99US-0146389Z
PR	03-AUG-1999	99US-0147004Z
PR	04-AUG-1999	99US-0147208Z

PR	04-AUG-1999:	99US-01471302:
PR	05-AUG-1999:	99US-01471260:
PR	05-AUG-1999:	99US-01471703P:
PR	06-AUG-1999:	99US-0147416P:
PR	06-AUG-1999:	99US-0147493P:
PR	09-AUG-1999:	99US-0147935P:
PR	10-AUG-1999:	99US-0148171P:
PR	11-AUG-1999:	99US-0148319P:
PR	12-AUG-1999:	99US-0148341P:
PR	13-AUG-1999:	99US-0148565P:
PR	13-AUG-1999:	99US-0148684P:
PR	16-AUG-1999:	99US-0149368P:
PR	17-AUG-1999:	99US-0149175P:
PR	18-AUG-1999:	99US-0149426P:
PR	20-AUG-1999:	99US-0149722P:
PR	20-AUG-1999:	99US-0149723P:
PR	20-AUG-1999:	99US-0149929P:
PR	22-AUG-1999:	99US-0149902P:
PR	22-AUG-1999:	99US-0149930D:
PR	25-AUG-1999:	99US-0150566P:
PR	26-AUG-1999:	99US-0150084P:
PR	27-AUG-1999:	99US-0151065P:
PR	27-AUG-1999:	99US-0151066P:
PR	27-AUG-1999:	99US-0151080P:
PR	30-AUG-1999:	99US-0151303P:
PR	31-AUG-1999:	99US-0151388P:
PR	01-SEP-1999:	99US-0151930P:
PR	07-SEP-1999:	99US-0152363P:
PR	10-SEP-1999:	99US-0153070P:
PR	11-SEP-1999:	99US-0153788P:
PR	12-SEP-1999:	99US-0154018P:
PR	16-SEP-1999:	99US-0154039P:
PR	20-SEP-1999:	99US-0154779P:
PR	22-SEP-1999:	99US-0155139P:
PR	22-SEP-1999:	99US-0155486P:
PR	24-SEP-1999:	99US-0155659P:
PR	28-SEP-1999:	99US-0156458P:
PR	29-SEP-1999:	99US-0156596P:
PR	04-OCT-1999:	99US-0157117P:
PR	05-OCT-1999:	99US-0157753P:
PR	06-OCT-1999:	99US-0157865P:
PR	07-OCT-1999:	99US-0158029P:
PR	08-OCT-1999:	99US-0158332P:
PR	12-OCT-1999:	99US-0158369P:
PR	13-OCT-1999:	99US-0159293P:
PR	13-OCT-1999:	99US-0159294P:
PR	13-OCT-1999:	99US-0159295P:
PR	14-OCT-1999:	99US-0159329P:
PR	14-OCT-1999:	99US-0159330P:
PR	14-OCT-1999:	99US-0159331P:
PR	14-OCT-1999:	99US-0159637P:
PR	14-OCT-1999:	99US-0159638P:
PR	18-OCT-1999:	99US-0159684P:
PR	21-OCT-1999:	99US-0160741P:
PR	21-OCT-1999:	99US-0160767P:
PR	21-OCT-1999:	99US-0160778P:
PR	21-OCT-1999:	99US-0160814P:
PR	21-OCT-1999:	99US-0160815P:
PR	22-OCT-1999:	99US-0160808P:
PR	22-OCT-1999:	99US-0160809P:
PR	22-OCT-1999:	99US-01608981P:
PR	26-OCT-1999:	99US-0160988P:
PR	26-OCT-1999:	99US-0161359P:
PR	26-OCT-1999:	99US-0161361P:
PR	28-OCT-1999:	99US-0161920P:
PR	28-OCT-1999:	99US-0161992P:
PR	29-OCT-1999:	99US-0161993P:
PR	29-OCT-1999:	99US-0162142P:





PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0158293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159333P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 23-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.

PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 29-OCT-1999; 99US-0162142P.  
  
Query Match 25.7%; Score 319; DB 3; Length 439;  
Best Local Similarity 36.2%; Pred. No. 1,5e-26;  
Matches 81; Conservative 31; Mismatches 82; Indels 30; Gaps 5;  
  
QY 1 LLAIVGVSDPFLILGSLDAPGDMLDKVARRLSLIKHPGCSMTSGSKALIEHLAKGNRF 60  
DB 224 LVLVLAHLKQLYQLGFTVDALIGAPDKTAKWGLDMH-----RSGGPAVEBLALBDPAK 278  
QY 61 HFDIKPEPLHAKKNDSPFTGLQHTVDKTIIMKKEKEEGEIKGQLLSA-----ADIAA 112  
DB 279 SVKENVPKTYHDCNCSYAGLKTQVRLAIEAK-----IDAKCPVSSATNEDRRARADIAA 334  
QY 113 TVQHTNACHLVKRTTHALIECKORDLLPONNAVLVASGVASNPYIRALEILTNAQTCT 172  
DB 335 SFQRAVALHLEKCEERADIDWALE--LEPSIKHVIISGVASNKYVRLRLNNIIVENLNK 391  
QY 173 LDCPPPRLLCTDNGSIMTANMGIETRLRAGLGILHIDIEGIRYEPKCP 216  
DB 392 LVCPPPSLCTDNGVMAWTGLLHFRVG-----RYDPPPP 425  
  
RESULT 13  
AA619285  
ID AAG19285 standard; protein; 444 AA.  
XX  
XX AAG19285;  
AC  
XX  
DT 17-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 21029.  
DE  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
XX Arabidopsis thaliana.  
PN  
XX  
XX EP1033405-A2.  
PD  
XX  
XX 06-SEP-2000.  
XX  
XX  
XX 25-FEB-2000; 2000EP-00301439.  
PF  
XX  
XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999;	99US-0134219P.
PR 14-MAY-1999;	99US-0134221P.
PR 14-MAY-1999;	99US-0134370P.
PR 18-MAY-1999;	99US-0134768P.
PR 19-MAY-1999;	99US-0134941P.
PR 20-MAY-1999;	99US-0135124P.
PR 21-MAY-1999;	99US-0135353P.
PR 24-MAY-1999;	99US-0135629P.
PR 25-MAY-1999;	99US-0136021P.
PR 27-MAY-1999;	99US-0136392P.
PR 28-MAY-1999;	99US-0136782P.
PR 01-JUN-1999;	99US-0137222P.
PR 03-JUN-1999;	99US-0137528P.
PR 04-JUN-1999;	99US-0137502P.
PR 07-JUN-1999;	99US-0137724P.
PR 08-JUN-1999;	99US-0138094P.
PR 10-JUN-1999;	99US-0138540P.
PR 10-JUN-1999;	99US-0138647P.
PR 14-JUN-1999;	99US-0139119P.
PR 16-JUN-1999;	99US-0139452P.
PR 17-JUN-1999;	99US-0139453P.
PR 17-JUN-1999;	99US-0139492P.
PR 18-JUN-1999;	99US-0139454P.
PR 18-JUN-1999;	99US-0139455P.
PR 18-JUN-1999;	99US-0139456P.
PR 18-JUN-1999;	99US-0139457P.
PR 18-JUN-1999;	99US-0139458P.
PR 18-JUN-1999;	99US-0139459P.
PR 18-JUN-1999;	99US-0139460P.
PR 18-JUN-1999;	99US-0139461P.
PR 18-JUN-1999;	99US-0139462P.
PR 18-JUN-1999;	99US-0139463P.
PR 18-JUN-1999;	99US-0139750P.
PR 18-JUN-1999;	99US-0139763P.
PR 21-JUN-1999;	99US-0139817P.
PR 22-JUN-1999;	99US-0139899P.
PR 23-JUN-1999;	99US-0140353P.
PR 23-JUN-1999;	99US-0140354P.
PR 24-JUN-1999;	99US-0140695P.
PR 28-JUN-1999;	99US-0140823P.
PR 29-JUN-1999;	99US-0140991P.
PR 30-JUN-1999;	99US-0141287P.
PR 01-JUL-1999;	99US-0141844P.
PR 01-JUL-1999;	99US-0142154P.
PR 02-JUL-1999;	99US-0142055P.
PR 06-JUL-1999;	99US-0142392P.
PR 08-JUL-1999;	99US-0142803P.
PR 09-JUL-1999;	99US-0142920P.
PR 12-JUL-1999;	99US-0142977P.
PR 13-JUL-1999;	99US-0143542P.
PR 14-JUL-1999;	99US-0143624P.
PR 15-JUL-1999;	99US-0144005P.
PR 16-JUL-1999;	99US-0144085P.
PR 16-JUL-1999;	99US-0144086P.
PR 19-JUL-1999;	99US-0144325P.
PR 19-JUL-1999;	99US-0144331P.
PR 19-JUL-1999;	99US-0144332P.
PR 19-JUL-1999;	99US-0144333P.
PR 19-JUL-1999;	99US-0144334P.
PR 19-JUL-1999;	99US-0144335P.
PR 20-JUL-1999;	99US-0144335P.
PR 20-JUL-1999;	99US-0144632P.
PR 20-JUL-1999;	99US-0144684P.
PR 21-JUL-1999;	99US-0144814P.
PR 21-JUL-1999;	99US-0145086P.
PR 21-JUL-1999;	99US-0145086P.
PR 22-JUL-1999;	99US-0145085P.
PR 22-JUL-1999;	99US-0145087P.
PR 22-JUL-1999;	99US-0145089P.
PR 22-JUL-1999;	99US-0145152P.
PR 23-JUL-1999;	99US-0145145P.
PR 23-JUL-1999;	99US-0145218P.
PR 23-JUL-1999;	99US-0145224P.
<hr/>	
PR 26-JUL-1999;	99US-0145276P.
PR 27-JUL-1999;	99US-0145913P.
PR 27-JUL-1999;	99US-0145918P.
PR 27-JUL-1999;	99US-0145919P.
PR 28-JUL-1999;	99US-0145951P.
PR 02-AUG-1999;	99US-0146386P.
PR 02-AUG-1999;	99US-0146388P.
PR 02-AUG-1999;	99US-0146389P.
PR 03-AUG-1999;	99US-0147038P.
PR 04-AUG-1999;	99US-0147204P.
PR 04-AUG-1999;	99US-0147302P.
PR 05-AUG-1999;	99US-0147192P.
PR 05-AUG-1999;	99US-0147260P.
PR 06-AUG-1999;	99US-0147303P.
PR 06-AUG-1999;	99US-0147416P.
PR 09-AUG-1999;	99US-0147493P.
PR 09-AUG-1999;	99US-0147935P.
PR 10-AUG-1999;	99US-0148171P.
PR 11-AUG-1999;	99US-0148319P.
PR 12-AUG-1999;	99US-0148341P.
PR 13-AUG-1999;	99US-0148655P.
PR 13-AUG-1999;	99US-0148684P.
PR 16-AUG-1999;	99US-0149368P.
PR 17-AUG-1999;	99US-0149175P.
PR 18-AUG-1999;	99US-0149426P.
PR 20-AUG-1999;	99US-0149722P.
PR 20-AUG-1999;	99US-0149723P.
PR 20-AUG-1999;	99US-0149929P.
PR 23-AUG-1999;	99US-0149902P.
PR 23-AUG-1999;	99US-0149930P.
PR 25-AUG-1999;	99US-0150566P.
PR 26-AUG-1999;	99US-0150884P.
PR 27-AUG-1999;	99US-0151065P.
PR 27-AUG-1999;	99US-0151066P.
PR 27-AUG-1999;	99US-0151080P.
PR 30-AUG-1999;	99US-0151303P.
PR 31-AUG-1999;	99US-015138P.
PR 01-SEP-1999;	99US-015130P.
PR 07-SEP-1999;	99US-0152363P.
PR 10-SEP-1999;	99US-0153070P.
PR 13-SEP-1999;	99US-0153758P.
PR 15-SEP-1999;	99US-0154018P.
PR 16-SEP-1999;	99US-0154039P.
PR 20-SEP-1999;	99US-0154779P.
PR 22-SEP-1999;	99US-0155139P.
PR 23-SEP-1999;	99US-0155486P.
PR 24-SEP-1999;	99US-0155659P.
PR 28-SEP-1999;	99US-0156458P.
PR 29-SEP-1999;	99US-0156596P.
PR 04-OCT-1999;	99US-0157117P.
PR 05-OCT-1999;	99US-0157753P.
PR 06-OCT-1999;	99US-0157865P.
PR 07-OCT-1999;	99US-0158029P.
PR 08-OCT-1999;	99US-0158232P.
PR 12-OCT-1999;	99US-0158369P.
PR 13-OCT-1999;	99US-0158923P.
PR 13-OCT-1999;	99US-0159294P.
PR 13-OCT-1999;	99US-0159295P.
PR 14-OCT-1999;	99US-0159329P.
PR 14-OCT-1999;	99US-0159330P.
PR 14-OCT-1999;	99US-0159331P.
PR 14-OCT-1999;	99US-0159637P.
PR 14-OCT-1999;	99US-0159638P.
PR 18-OCT-1999;	99US-0159584P.
PR 21-OCT-1999;	99US-0160741P.
PR 21-OCT-1999;	99US-0160767P.
PR 21-OCT-1999;	99US-0160768P.
PR 21-OCT-1999;	99US-0160770P.
PR 21-OCT-1999;	99US-0160814P.
PR 21-OCT-1999;	99US-0160815P.
PR 22-OCT-1999;	99US-0160980P.
PR 22-OCT-1999;	99US-0160981P.
PR 22-OCT-1999;	99US-0160989P.

PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161921P.  
PR 28-OCT-1999; 99US-0161933P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 25.7%; Score 319; DB 3; Length 444;  
Best Local Similarity 36.2%; Pred. No. 1.5e-26;  
Matches 81; Conservative 31; Mismatches 82; Indels 30; Gaps 5;

QY 1 LIALVGVSDFLILGKSLDIPGMDLKVARRLSLIKPECTSGGKAIETHLAKQGNRF 60  
DB 229 LVLIAHKKIGQYTOGTVDDAIGFAFDYAKWGLDMH-----RSGGPAVEELALEGDAK 283  
QY 61 HFDIKPLHAHAKNCDSPFTGLQHVVDKTIIMKKEEGIEKGQILSSA-----ADIAA 112  
DB 284 SVKRNVPKXHKDCNFSYAGLKTQVRLAIEAKE---IDAKCPVSATMEDRRRADIAA 339  
QY 113 TVQHTMACHLVKRTTHRAILFCQKRDLPQNNNAVVASGVASNFYIRRALETLTATQCT 172  
DB 340 SFQVAVVHLHEKCKERAILDWALE--LEPSIKHWISGVASNKCYRLANIVENKOLX 396  
QY 173 LLCPPRLCTNGMIAMNGIERLRAGIGILHDIGIRYEPKCP 216  
DB 397 LVCPSPSLCTDNGVAVMTGLEHFRVG-----RYDPPPP 430

## RESULT 14

ID ABU22934 standard; protein; 382 AA.

AC ABU22934;

DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #8461.

KM Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Bordetella pertussis.

PN MO200277183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR N-PSDB; ACA26804.

XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
PS Claim 25; SEQ ID NO 50858; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
SQ Sequence 382 AA;

Query Match 20.7%; Score 257; DB 6; Length 382;  
Best Local Similarity 36.1%; Pred. No. 1.2e-19;  
Matches 74; Conservative 33; Mismatches 78; Indels 20; Gaps 7;

QY 2 LALVGVSDFLILGKSLDIPGMDLKVARRLSLIKPECTSGGKAIETHLAKQGNRF 61

DB 178 LMLVGVGRYELLETLDNAGEAPKSAKMLG-CYP-----GCPALARLAEQDASR 230

QY 62 FDIKPLHAHAKNCDSPFTGLQ-HVTDKTIIMKKEEGIEKGQILSSAADIAATVQHTMAC 120

DB 231 YDLPRFMLSGLDFFSGLKTAVLTRV--KATRDGGEIGR--QBRADIAATQAAIVE 286

QY 121 HLVKRTTHRAILFCQKRDLPQNNNAVVASGVASNFYIRRALETLTATQCTLLCPPRR 180

DB 287 VLAAKAIRAL--KQTGL-----RRLVAVAGVGANLRLRALKPLRAEAYFPPLSL 338

QY 181 CTDNGMIMAMNGIERLRAGIGILHD 205

DB 339 CTDNGMIMAMNGIERLRAGIGILHD 363

## RESULT 15

ID ABU35232 standard; protein; 348 AA.

AC ABU35232;

DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #20759.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Moraxella catarrhalis.

OS WO200277183-A2.

XX 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW,  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR N-PSDB; ACA39102.  
XX

PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 63156; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation; (7) identifying a gene in an operon required for  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_ptc\_sequences  
XX

SQ Sequence 348 AA;

Query Match 20.0%; Score 248.5; DB 6; Length 348;  
Best Local Similarity 33.2%; Pred. No. 9.4e-19;  
Matches 66; Conservative 33; Mismatches 79; Indels 21; Gaps 5;

QY 1 LLAIVGVSPDLAKSLIDIAAGDMLDKVARRSLIKAPBCSTWSGKAIEHLAKQNR 60  
DB 147 MLVRADGVGVYQIIGESIDDAVGECPDKTAKMLK-LYP-----GPNIEKLAKNPNH 199  
QY 61 HFDIKPPLHAKNCDPSTFTGLQHTVDKLIKKEKEEGIEKGQIISAAADIAATVOHTMAC 120  
DB 200 AYELEPRPMOH-KGLDFSPSGKTKTAIHNLIKDTPNQSPP-----ATRADIAASFVAVD 253  
QY 121 HLVRGTRALIFCQRODLPGNNNAVIVSGVASFYIRALBITNATQCTILCPPRL 180  
DB 254 TLVKKCTVALGPTGCIHQ-----LVVAGSVSANQPLRKILVETLRQIDASVITAPTEL 305  
QY 181 CTDNGIMIAMNGIERLRAG 199  
||||| ||| | | |

DB 306 CTDNGAMIAVAGFCRLSCG 324  
Search completed: November 10, 2005, 16:48:03  
Job time : 65.4261 secs



C;Accession: AB2902  
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kueyavin, T.; Levy, R.; Li, M.; McClellan  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krepsan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A;Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
A;Accession: AB2577; MUID:21608550; PMID:11743193  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-365 <KUR>  
A;Cross-references: GB:AB006688; PIDN:AAL43632.1; PID:g1741154; GSPDB:GN00186  
A;Experimental source: strain C58 (Dupont)  
C;Genetics:  
A;Gene: gcp  
A;Map position: circular chromosome  
C;Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.7%; Score 244; DB 2; Length 365;  
Best Local Similarity 31.9%; Pred. No. 1.1e-14;  
Matches 73; Conservative 34; Mismatches 92; Indels 30; Gaps 7;

Qy 2 LALVGVSDPFLILGKSLDIPAGDMDKVARRSLIKHPCSTWGGKALEHLAKOGRFH 61  
Db LVLVRGVEYERWGTITDIALGEAFDPTAKLGL-PPY-----GSPAVENAAAGDPDR 201  
149 LVLVRGVEYERWGTITDIALGEAFDPTAKLGL-PPY-----GSPAVENAAAGDPDR 201

Qy 62 FDIKPEPLHAKNCDPFSFTGL-----QHTVDKIIMKEKEGIEKGQILSSADIAATVQHT 117  
Db FPLRPFWGGEARLDPSFGSLKTAVAQAATAIAPLSEOD-----IADICASFQRA 250

Qy 118 MACHLVKTRTHAILEFCRKDRLPQNN--VVASGVSANFYIRALLETITNATQCTLLC 175  
Db VSRILKRIKIGRLARFKE--PPHINSEPALVAVAGVANOIRQLTQLCTHGRFVA 308

Qy 176 PPRRLCTDNGIMIANWGIERLAGLILHDIEGIRYEPKCPGLVDSKE 224  
Db PPHRLCTDNGIMIANWGIERLAGLILHDIEGIRYEPKCPGLVDSKE 353

RESULT 3  
D97677  
probable o-sialoglycoprotein endopeptidase (glycoproteinase) (imported) - *Agrobacterium*  
C;Species: *Agrobacterium tumefaciens*  
C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C;Accession: D97677  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tum*  
A;Reference number: A97359; MUID:21608551; PMID:11743194  
A;Accession: D97677  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-366 <KUR>  
A;Cross-references: GB:AB007869; PIDN:AAK88373.1; PID:g15157858; GSPDB:GN00169  
C;Genetics:  
A;Gene: AGR\_C\_4806  
A;Map position: circular chromosome  
C;Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.7%; Score 244; DB 2; Length 366;  
Best Local Similarity 31.9%; Pred. No. 1.1e-14;  
Matches 73; Conservative 34; Mismatches 92; Indels 30; Gaps 7;

Qy 2 LALVGVSDPFLILGKSLDIPAGDMDKVARRSLIKHPCSTWGGKALEHLAKOGRFH 61  
Db LVLVRGVEYERWGTITDIALGEAFDPTAKLGL-PPY-----GSPAVENAAAGDPDR 202  
150 LVLVRGVEYERWGTITDIALGEAFDPTAKLGL-PPY-----GSPAVENAAAGDPDR 202

Qy 62 FDIKPEPLHAKNCDPFSFTGL-----QHTVDKIIMKEKEGIEKGQILSSADIAATVQHT 117  
Db FPLRPFWGGEARLDPSFGSLKTAVAQAATAIAPLSEOD-----IADICASFQRA 251

Qy 118 MACHLVKTRTHAILEFCRKDRLPQNN--VVASGVSANFYIRALLETITNATQCTLLC 175  
Db VSRILKRIKIGRLARFKE--PPHINSEPALVAVAGVANOIRQLTQLCTHGRFVA 309

Qy 176 PPRRLCTDNGIMIANWGIERLAGLILHDIEGIRYEPKCPGLVDSKE 224  
Db PPHRLCTDNGIMIANWGIERLAGLILHDIEGIRYEPKCPGLVDSKE 354

RESULT 4  
AB3274  
O-sialoglycoprotein endopeptidase (BC 3.4.24.57) (imported) - *Brucella melitensis* (strai  
C;Species: *Brucella melitensis*  
C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 15-Feb-2002  
C;Accession: AB3274  
R;DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Ios, T.; Ivanova,  
.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A;Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
A;Reference number: AD3252; PMID:11756688  
A;Accession: AB3274  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-359 <KUR>  
A;Cross-references: GB:AB008917; PIDN:AAL51357.1; PID:g17982056; GSPDB:GN00190  
A;Experimental source: strain 16M  
C;Genetics:  
A;Gene: BME10175  
A;Map position: I  
C;Superfamily: O-sialoglycoprotein endopeptidase  
C;Keywords: hydrolase; metalloproteinase

Query Match 19.5%; Score 241.5; DB 2; Length 359;  
Best Local Similarity 34.5%; Pred. No. 1.9e-14;  
Matches 69; Conservative 27; Mismatches 77; Indels 27; Gaps 6;

Qy 2 LALVGVSDPFLILGKSLDIPAGDMDKVARRSLIKHPCSTWGGKALEHLAKOGRFH 61  
Db LVLVRGVEYERWGTITDIALGEAFDPTAKLGL-PPY-----GSPAVENAAAGDPDR 197

Qy 62 FDIKPEPLHAKNCDPFSFTGL-----QHTVDKIIMKEKEGIEKGQILSSADIAATVQHT 116  
Db FPLRPFWGGEARLDPSFGSLKTAVAQAATAIAPLSEOD-----IADICASFQRA 245

Qy 117 MACHLVKTRTHAILEFCRKDRLPQ--NNAVVASGVSANFYIRALLETITNATQCTLLC 175  
Db AVADTSLDSVGRSLERFKE--FPDCATPSLVAVAGVANKTIRALLETITNATQCTLLC 303

Qy 176 PPRRLCTDNGIMIANWGIER 195  
Db PPHRLCTDNGIMIANWGIER 323

RESULT 5  
E71711  
probable o-sialoglycoprotein endopeptidase (gcp) RP037 - *Rickettsia prowazekii*  
C;Species: *Rickettsia prowazekii*  
C;Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 09-Jul-2004  
C;Accession: E71711  
R;Andersson, S.G.E.; Zomrodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U.  
Nature 396, 133-140, 1998  
A;Title: The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria.  
A;Reference number: A71630; MUID:99039499; PMID:9823893  
A;Accession: E71711  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-387 <AND>  
A;Cross-references: UNIPROT:Q9ZEA8; GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA1450  
A;Experimental source: strain Madrid E  
C;Genetics:  
A;Gene: gcp; RP037  
C;Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.2%; Score 238; DB 2; Length 387;  
Best Local Similarity 23.7%; Pred. No. 4.4e-14;  
Matches 70; Conservative 32; Mismatches 74; Indels 60; Gaps 6;

5 VGVSDPFLLLGKSLDIAPGMDLVKVARRLSLIKHPECSTMSGKALIEHLAKGNRHPDI 64  
145 VLGIKTKYKILGTTIDDAVGEFEDKVAKMLNTL-----SFGGPEIEKRAKLGPHKXKF 197  
65 KRPPLHAKKNDPSTFGLOHTVDKIIIMK-KEKEEGIEKGQILSSAADIAATVOHTMACHLV 123  
198 KPPIINSNCNMSFSGLKTAVRTLIIMLKENVDSV-----INDIASPFTIGALLS 249

124 KRTTRAILFCQKQ-----RD-----LLPON----- 142  
250 SKMDAIRLYKQIINDYEDINHPKLNLSKSPKDEPNMKPLECTRPKYRIHIQNSYRS 309

143 ---NAVVASGVSANFYIRALIELTNATQCTLLCPPRILCTDNGMIAMNGIER 195  
310 NLINDTIVIAGVANAKYLOEILSDCTRPYGRILIAPMHLCIDNAAAMIAVAGLER 365

RESULT 6  
F87257  
peptidase M22 family protein [imported] - Caulobacter crescentus  
C:/Species: Caulobacter crescentus  
C:/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:/Accession: F87257

R;Nieman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Laub, M.T.; Deboy, R.T.; Dodson, R.U.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kjoltor n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4133-4141, 2001

A:/Title: Complete Genome Sequence of Caulobacter crescentus.  
A:/Reference number: A87249; WUID:21173698; PMID:11259647  
A:/Accession: F87257  
A:/Status: preliminary  
A:/Molecule type: DNA  
A:/Residues: 1-367 <STO>  
A:/Cross-references: GB:AE005673; NID:g13421168; PIDN:AAK22058.1; GSPDB:GN00148  
C:/Genetics:

A:/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.2%; Score 237.5; DB 2; Length 367;  
Best Local Similarity 34.3%; Pred. No. 4.6e-14;  
Matches 73; Conservative 29; Mismatches 90; Indels 21; Gaps 6;

5 VGVSDPFLLLGKSLDIAPGMDLVKVARRLSLIKHPECSTMSGKALIEHLAKGNRHPDI 64  
156 VSGVGACRRLGTTIDDAAGEAFDKIAKSLGL-PPY-----GCPALERTAVGDPTRYAL 208

65 KRPPLHAKKNDPSTFGLOHTVDKIIIMKKEKEEGIEKGQILSSAADIAATVOHTMACHLV 124  
209 PRALLGRKDCDFSSGKTAARJAETLTTDD-----ARBDLAAGVAPAAIARQISE 259

125 RTRRAILFCQKQKRDLPQNNNAVVASGVANFYIRALIELTNATQCTLLCPPRILCTDN 184  
260 RVRBAKLYK--DSHDEDLAFVAVAGVANGAVAAALADCEKNGFSFAAPLAYCTDN 317

185 GIMIAMNGIERLRAGLILHDIEGIRYEPKPL 217  
318 AAMIALAGARL--ALGIFDDLDIAI-RPRMPL 347

RESULT 7  
H83572  
O-sialoglycoprotein endopeptidase PA0580 [imported] - Pseudomonas aeruginosa (strain PAO C:/Species: Pseudomonas aeruginosa  
C:/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:/Accession: H83572

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidig, K.; Lim, .; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:/Reference number: A82950; WUID:20437337; PMID:10984043  
A:/Accession: H83572  
A:/Status: preliminary  
A:/Molecule type: DNA  
A:/Residues: 1-341 <STO>  
A:/Cross-references: GB:AE004494; GB:AE004091; NID:g9946446; PIDN:AAQ03969.1; GSPDB:GN001  
A:/Experimental source: strain PA01  
C:/Genetics:

A:/Gene: gcp; PA0580  
C:/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 18.6%; Score 231; DB 2; Length 344;  
Best Local Similarity 33.8%; Pred. No. 1.6e-13;  
Matches 73; Conservative 34; Mismatches 87; Indels 22; Gaps 8;

2 LALVGVSDPFLLLGKSLDIAPGMDLVKVARRLSLIKHPECSTMSGKALIEHLAKGNRHP 61  
142 LVRVIGRIGRYOLLGSDVDAAGEAFDKIAKSLGL-GYP-----GCPETARILAEKGTGPR 194

62 FDIKRPPLHAKKNDPSTFGLOHTVDKIIIMKKEKEEGIEKGQILSSAADIAATVOHTMACH 121  
195 FVFRPMTDRPGLDPSFSGLKTFTLN-TWQCVKAGDDBEQ---TRCDIALAFQTAVERT 250

122 LVKRTTRAILFCQKQKRDLPQNNNAVVASGVANFYIRALIELTNATQCTLLCPPRILC 181  
251 LIIKCRRL--KQGL--KN---LVYAGVSANQALRSGLEKMGEMKGVYVYARPRC 302

182 TDNGMIAMNGIERLRAGLILHDIEGIRYEPKPL 217  
303 TDNGAMIAVAGCORLLAG---CHDGPALISVOPRPMW 335

RESULT 8  
E97707  
O-sialoglycoprotein endopeptidase (BC 3.4.24.57) [imported] - Rickettsia conorii (strain C:/Species: Rickettsia conorii  
C:/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 22-Oct-2001  
C:/Accession: E97707

R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro Science 293, 2093-2098, 2001

A:/Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A:/Reference number: A97700; WUID:21442074; PMID:11557893  
A:/Accession: E97707  
A:/Status: preliminary  
A:/Molecule type: DNA  
A:/Residues: 1-344 <KUR>  
A:/Cross-references: GB:AE006914; PIDN:AAL02599.1; PID:g15619097; GSPDB:GN00173  
C:/Genetics: gcp  
C:/Superfamily: O-sialoglycoprotein endopeptidase  
C:/Keywords: hydrolase; metalloproteinase

Query Match 18.6%; Score 230.5; DB 2; Length 344;  
Best Local Similarity 33.2%; Pred. No. 1.9e-13;  
Matches 64; Conservative 34; Mismatches 76; Indels 19; Gaps 5;

5 VGVSDPFLLLGKSLDIAPGMDLVKVARRLSLIKHPECSTMSGKALIEHLAKGNRHPDI 64  
145 VLGIKTKYKILGTTIDDAVGEAFDKVAKMLNTL-----AFPGPEIEKRAKLGDPHKXKF 197

65 KRPPLHAKKNDPSTFGLOHTVDKIIIMK-KEKEEGIEKGQILSSAADIAATVOHTMACHLV 123  
198 KPPIINSNCNMSFSGLKTAVRTLIIMLKENVDSV-----INDIASPFTIGALLS 249

124 KRTTRAILFCQKQKRDLPQNNNAVVASGVANFYIRALIELTNATQCTLLCPPRILCT 182  
250 SKMDAIRLYKQIINDYEDINHPKLNLSKSPKDEPNMKPLECTRPKYRIHIQNSYRS 307

183 DNGIAMIAMNGIER 195  
308 DNAAAMIAVAGLER 320



## RESULT 9

H64074

O-sialoglycoprotein endopeptidase (EC 3.4.24.57) - Haemophilus influenzae (strain Rd KW2)  
N/Alternate names: sialoglycoproteinase  
C/Species: Haemophilus influenzae  
C/Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 09-Jul-2004  
C/Accession: H64074

R./Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kesteven, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.D.M.; Brandon, R.C.; FINE, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M. Science 269, 496-512, 1995

A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A./Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A/Reference number: A64000; MUID:95350630; PMID:7542800

A/Accession: H64074  
A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA  
A/Residues: 1-342 <TIGR>

A/Cross-references: UNIPROT:P43764; GB:U32735; GB:I42023; NID:G1573509; PIDN:AAC22187.1; C/Superfamily: O-sialoglycoprotein endopeptidase  
C/Keywords: hydrolase, metalloproteinase

Query Match 18.1%; Score 224; DB 2; Length 342;  
Best Local Similarity 33.5%; Pred. No. 7.3e-13;  
Matches 67; Conservative 28; Mismatches 83; Indels 22; Gaps 6;

QY 2 LALVGVSDFLLLGKSLDIAPGMDLKVARRLSLTKHPECSTMSGGKAIEHLAKQGNRFH 61

Db 142 LVKVDGVGVKEVISESIDDAAGEAFDKTKLGL-DYP-----GGALISRLAEKGTNR 194

QY 62 FDIKPLHNAKNCPSFTGLQHTVDKIMKKEKESG--IEKQIILSSAADIATVQHTMA 119

Db 195 FTFPRPMTDRAGLPFGSLKTPAANTVQAATKNGELLEQ----TKADIVARQDAVV 249

QY 120 CHLVKTRTRALIFCKQRDLTPONNAVLVAGGVASNFYIRRALIEITLTAQCTLLCPPR 179

Db 250 DTLA-----IKCK-RALKETGYRLVIAGSVSNKKRLRETLHLMLQGLGEVYVPRQ 301

QY 180 LCTDNGMIANNGIERLRAG 199

Db 302 FCTDNGAMIAVTGFLRLKQG 321

## RESULT 10

G69388

O-sialoglycoprotein endopeptidase homolog - Archaeoglobus fulgidus  
C/Species: Archaeoglobus fulgidus  
C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 21-Jul-2000  
C/Accession: G69388

R./Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L. Nature 390, 364-370, 1997

A/Authors: Uitterlinden, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S. Smith, H.O.; Moese, C.R.; Venter, J.C.  
A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo

A/Reference number: A69250; MUID:98049343; PMID:9389475

A/Accession: G69388  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA  
A/Residues: 1-323 <KLE>

A/Cross-references: GB:AE001027; GB:AE000782; NID:G2689350; PIDN:AAB90129.1; PID:G264947  
C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 17.9%; Score 221.5; DB 2; Length 323;  
Best Local Similarity 32.8%; Pred. No. 1.2e-12;  
Matches 62; Conservative 30; Mismatches 66; Indels 31; Gaps 8;

QY 11 FLILGKSLDIAPGMDLKVARRLSLTKHPECSTMSGGKAIEHLAKQGNRFHDIKPLH 70

Db 142 YRVGERTLIDIGNALDLKLRHMG-LKP-----GGRKIEELAKKGKYNH--LP--YV 190

QY 71 AKNCDFSTGLQHTVDKIMKKEKESGIEKQIILSSAADIATVQHTMACHLVKTRTRAI 130

Db 191 VKGMDPSFSGVTAQORLP-----DSGVV-----MEDVARSFQETAPAMLTREVERAL 238

QY 131 LFCKQRDLTPONNAVLVAGGVASNFYIRRALIEITLTAQCTLLCPPRLLCTDNGIMIM 190

Db 239 AV-----LDLNEVLIV--GGVANKRQLQEMLRIMCEDRGAKFYVPKELAGDNGAMIAV 290

QY 191 NGIERLRAG 199

Db 291 TGLMTKKG 299

## RESULT 11

AG0892

probable glycoprotease (imported) - Salmonella enterica subsp. enterica serovar Typhi (s)  
C/Species: Salmonella enterica subsp. enterica serovar Typhi  
A/Note: this species has also been called Salmonella typhi  
C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C/Accession: AG0892

R./Parhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Comerston, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P. Nature 413, 848-852, 2001

A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A/Reference number: AB0502; MUID:21534947; PMID:11677608

A/Accession: AG0892  
A/Status: preliminary  
A/Molecule type: DNA

A/Residues: 1-337 <PAR>  
A/Cross-references: GB:AL513382; PIDN:CAD07733.1; PID:G16504285; GSPDB:GN00176

C/Accession: STY3387  
C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 17.2%; Score 213; DB 2; Length 337;  
Best Local Similarity 30.6%; Pred. No. 7.4e-12;  
Matches 66; Conservative 32; Mismatches 92; Indels 26; Gaps 7;

QY 2 LALVGVSDFLLLGKSLDIAPGMDLKVARRLSLTKHPECSTMSGGKAIEHLAKQGNRFH 61

Db 142 LISVTGIGGYELLGESSIDDAAGEAFDKTKLGL-DYP-----GGPMLSKMASQGTAGR 194

QY 62 FDIKPLHNAKNCPSFTGLQHTVDKIMKKEKESGIEKQIILSSAADIATVQHTMA 121

Db 195 FTFPRPMTDRPGLDPSFGSLKTPAANTIRSGDDE-----QTRADIARAFEDAVDT 246

QY 122 LVKTRTRALIFCKQRDLTPONNAVLVAGGVASNFYIRRALIEITLTAQCTLLCPPRLC 181

Db 247 L-----MIKCK-RALBESTGFRGLVMAGSVSNKRLRLALALAMQGRREYVYARPERC 298

QY 182 TDNGMIANNGIERLRAGGLIHDIGIRYKPCPL 217

Db 299 TDNGAMIAVAGVRFKA--GVTA DL-GVTVRPRWPL 331

## RESULT 12

G70369  
sialoglycoproteinase - Aquifex aeolicus

C/Species: Aquifex aeolicus  
C/Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004  
C/Accession: G70369

R./Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O. V. Nature 392, 353-358, 1998

A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A/Reference number: A70300; MUID:98196666; PMID:9537320

A/Accession: G70369  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA  
A/Residues: 1-335 <AQF>

A:Cross-references: UNIPROT:066986; GB:AE000708; NID:g2983356; PIDN:AAC06951.1; PID:g298  
A:Experimental source: strain V85  
C:Genetics:  
A:Gene: gcp  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 17.0%; Score 210.5; DB 2; Length 335;  
Best Local Similarity 32.0%; Pred. No. 1.2e-11;  
Matches 70; Conservative 41; Mismatches 73; Indels 35; Gaps 10;

QY 2 LALVGVSDPFLILGSLDIAPGMDLKVARRLSLKHPECSMTSGSKAIEHLAKOGRPH 61  
DB 141 LYLVRDPRYDFLGGTLDVAAGEAYDKVAKMLGL-GYF-----GGPIIDRLAKESKKU- 192  
QY 62 FDIKPEPLHAKNCDFSTGLQHVTDKIIIMKEKEGIEKGQILSSAADIATVQHTMAC 121  
DB 193 YPLKPEPLMEEGNLNFSGLK---TALNLLKKEKNVR-----EDLIASFQETVVEI 242  
QY 122 LVKRTTRAILPCKQDILLPQNNNAVIVASGVASNFYIRALBITLNATQ---CTLLCPPP 178  
DB 243 LLEKS---LWAMKKTGIRK---LVVVGVSANSRLR---EVFKASQEGYFELYIPIH 291  
QY 179 RLCTDNGIMIANGIERLRAGLGLHDIEGIRYEPKCP 217  
DB 292 SLSTDNALMIAYAGMERFRKGVADLVNP--QENIPL 327

## RESULT 13

AI0079  
probable glycoproteinase gcp [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis  
C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001

C:Accession: AI0079  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tiltball, R.W.; Holden, M.T.G.; Prentice, M.B.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11563360

A:Accession: AI0079  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-337 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC89500.1; PID:g15978736; GSPDB:GN00175  
C:Genetics:  
A:Gene: gcp  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 16.9%; Score 210; DB 2; Length 337;  
Best Local Similarity 30.0%; Pred. No. 1.4e-11;  
Matches 65; Conservative 33; Mismatches 91; Indels 28; Gaps 6;

QY 2 LALVGVSDPFLILGSLDIAPGMDLKVARRLSLKHPECSMTSGSKAIEHLAKOGRPH 61  
DB 142 LLSVTGIGYEYLLIGESVDADAGEAFDXTAKLLGL-DYP-----GSPMLSRMAOQTVGR 194  
QY 62 FDIKPEPLHAKNCDFSTGLQHVTDKIIIMKEKEGIEKGQILSSAADIATVQHTMAC 121  
DB 195 FTFPRPMTDRPGLDFSTGLKTFPAANTIRANGDD-----QTRADIARAFEDVVDVT 246  
QY 122 LVKRTTRAILPCKQDILLPQNNNAVIVASGVASNFYIRALBITLNATQCTLLCPPPL 180  
DB 247 LAISKKA-----LDQGFKRLVITAGVSAQOTLRLLKADMMQKRGSGVFPARPER 297  
QY 181 CTDNGIMIANGIERLRAGLGLHDIEGIRYEPKCP 217  
DB 298 CTDNGAMIAVAGVRLRGN---LNGELSVSRPRPL 331

RESULT 14  
F75029  
O-sialoglycoprotein endopeptidase (gcp) PAB1159 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi

C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004  
C:Accession: F75029  
R:anonymous, Genoscope  
A:Submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru

A:Reference number: A75001  
A:Accession: F75029  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-324 <KAM>  
A:Cross-references: UNIPROT:Q9UXK7; GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CA506

A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: gcp; PAB1159  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 16.9%; Score 209; DB 2; Length 324;  
Best Local Similarity 32.5%; Pred. No. 1.6e-11;  
Matches 65; Conservative 34; Mismatches 69; Indels 32; Gaps 8;

QY 1 LALVGVSDPFLILGSLDIAPGMDLKVARRLSLKHPECSMTSGSKAIEHLAKOGRPH 60  
DB 135 VLAIEGG--RYRVFGTLDIGNALDVPAREGL-----GPGGPKVEKLAKEGY 185  
QY 61 HDIKPEPLHAKNCDFSTGLQHVTDKIIIMKEKEGIEKGQILSSAADIATVQHTMAC 120  
DB 186 ---IELP-YAVKGMDSFGL--LTEAIRKYSRGKRV-----DLAYSFOETAF 230  
QY 121 LVKRTTRAILPCKQDILLPQNNNAVIVASGVASNFYIRALBITLNATQCTLLCPPPL 180  
DB 231 ALVEYTERAVANTE-----KQEVVIV--GGVANNRLRMLRIMTEDRGIKFVPPYDL 282  
QY 181 CTDNGIMIANGIERLRAGL 200  
DB 283 CRDNGAMIAVYTGRLKAGI 302

## RESULT 15

C91122

probable O-sialoglycoprotein endopeptidase [imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: C91122  
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasekawa, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C91122  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-337 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA037370.1; PID:g13363420; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain R1MD 0509952

C:Genetics:  
A:Gene: EC9347  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 16.9%; Score 209; DB 2; Length 337;  
Best Local Similarity 30.6%; Pred. No. 1.7e-11;  
Matches 66; Conservative 34; Mismatches 90; Indels 26; Gaps 8;

QY 2 LALVGVSDPFLILGSLDIAPGMDLKVARRLSLKHPECSMTSGSKAIEHLAKOGRPH 61  
DB 142 LLSVTGIGYEYLLIGESVDADAGEAFDXTAKLLGL-DYP-----GSPMLSRMAOQTVGR 194  
QY 62 FDIKPEPLHAKNCDFSTGLQHVTDKIIIMKEKEGIEKGQILSSAADIATVQHTMAC 121  
DB 195 FTFPRPMTDRPGLDFSTGLKTFPAANTIRANGDD-----QTRADIARAFEDVVDVT 246  
QY 122 LVKRTTRAILPCKQDILLPQNNNAVIVASGVASNFYIRALBITLNATQCTLLCPPPL 181  
DB 247 LAIKCKRAL-----DLNFKR--LVMAAGVSAHNRILRAKLAEMMKRGSGVFPARPER 298

Mon Nov 14 07:47:35 2005

us-10-649-273-2\_copy\_176\_414.rpr

**Page 6**

```

Qy      182 TDNGIMAMNGIERLRAGGILHDIEGIRYEPKCP 217
          ||| ||| : : | : : ||
Db      299 TDNGAMIAVAGMVFKA--GATADL-GVSVREPRWPL 331

```

Search completed: November 10, 2005, 16:53:09  
Job time : 15.8076 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: November 10, 2005, 15:16:01 ; Search time 57.9315 Seconds  
(without alignments)  
2112.614 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 LLLLVGVSDFLILGKSLDI.....DISKEVGASIKVQLKKEI 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_03:.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1240	100.0	096EV9	O96EV9 homo sapien
2	1213.5	97.9	094B0	O94B0 homo sapien
3	1086	87.6	06PEB4	O6PEB4 mus musculu
4	1079	87.0	08BLB6	O8BLB6 mus musculu
5	1071	86.4	09D0N0	O9D0N0 mus musculu
6	983	79.3	06AYN7	O6AYN7 rattus norv
7	960	77.4	096NH5	O96NH5 homo sapien
8	751	60.6	08JFW3	O8JFW3 brachydanto
9	747	60.2	08JFR7	O8JFR7 brachydanto
10	361.5	29.2	07Q9I8	O7Q9I8 anopheles g
11	332.5	26.8	09VMD6	O9VMD6 drosophila
12	332.5	26.8	09VMD6	O9VMD6 drosophila
13	324	26.1	022145	O22145 arabidopsis
14	299.5	24.2	0677H2	O677H2 hyacinthus
15	262.5	21.2	092LH8	O92LH8 rhizobium m
16	257	20.7	07VXN4	O7VXN4 bordetella
17	254	20.5	073H71	O73H71 wolbachia p
18	253	20.4	098E16	O98E16 rhizobium p
19	252.5	20.4	06FCR3	O6FCR3 acinetobact
20	252	20.3	07M6E8	O7M6E8 bordetella
21	251	20.2	07M134	O7M134 bordetella
22	246.5	19.9	06NDX3	O6NDX3 rhodospseudo
23	245	19.8	08UC47	O8UC47 agrobacteri
24	244	19.7	08UC47	O8UC47 agrobacteri
25	244	19.7	07CWJ8	O7CWJ8 agrobacteri
26	241.5	19.5	08YJBI	O8YJBI bruceella me
27	241.5	19.5	08FYI5	O8FYI5 bruceella su
28	241	19.3	09CLJ1	O9CLJ1 pasteurella
29	239	19.3	093FU2	O93FU2 cowdria rum
30	238	19.2	09E2A8	O9E2A8 rickettsia
31	237.5	19.2	09ABZ9	O9ABZ9 caulobacter

32	236.5	19.1	357	2	089MW1	O89MW1 bradyrhizob
33	234.5	18.9	344	2	07PAG7	O7PAG7 rickettsia
34	233.5	18.8	339	2	06LV10	O6LV10 photobacter
35	231	18.6	341	2	07NUE3	O7NUE3 chromobacter
36	231	18.6	341	2	091SV7	O91SV7 pseudomonas
37	231	18.6	364	2	06G1R3	O6G1R3 bartonella
38	230.5	18.6	344	2	09ZAK6	O9ZAK6 rickettsia
39	228.5	18.4	353	2	07VQO9	O7VQO9 candidatus
40	224	18.1	337	2	082XN2	O82XN2 nitrosomonas
41	224	18.1	342	1	GCP HABIN	P43764 haemophilus
42	223.5	18.0	344	2	06SRP0	O6SRP0 manheimia
43	221.5	17.9	323	2	0291S3	O291S3 archaeglob
44	219.5	17.7	347	2	08XX97	O8XX97 ralestonia s
45	217	17.5	341	2	08BS16	O8BS16 oceanobact1

## ALIGNMENTS

RESULT 1  
O96EV9 PRELIMINARY; PRT; 414 AA.  
ID O96EV9  
AC O96EV9; 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
DE OSGEPL1 protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Muscle;  
RX MEDLINE=22388257; PubMed12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schaller G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,  
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richardson S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
FAhey J., Helton B., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Bouffard G.G.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smalios D.E., Scherch A., Schein J.B.,  
RA Jones S.J., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Muscle;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; BC011904; AAH1904.1; -  
DR MEROPS; M22.004; -  
DR GO; GO:0008450; F-0-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008270; F-0-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
DR Pfam; PF00814; Peptidase\_M22\_1.  
DR PIRSF; PIRSF004537; Osialgl\_ptide; 1.  
DR PRINTS; PR00789; OSIALOPTASE.  
DR Prodom; PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs; TIGR00329; GCP\_1\_  
SQ SEQUENCE 414 AA; 45122 MW; A536B333F5C6B8DD CRC64;

Query Match 100.0%; Score 1240; DB 2; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 5,6e-101;  
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLAIVGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSMTSGGKAIIEHLAKQGNRF 60  
 DB 176 LLAIVGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSMTSGGKAIIEHLAKQGNRF 235  
 QY 61 HFDIKPPLHAKNCDPSFTGLOHVTDKIIMKKEKEGIEKGQILSSAADIAATVOHTMAC 120  
 DB 236 HFDIKPPLHAKNCDPSFTGLOHVTDKIIMKKEKEGIEKGQILSSAADIAATVOHTMAC 295  
 QY 121 HLVRKTRHAILFCCKORDLLPQNNAVLVASGVASFYIRRALEILTNAQTCTLLCPPRL 180  
 DB 296 HLVRKTRHAILFCCKORDLLPQNNAVLVASGVASFYIRRALEILTNAQTCTLLCPPRL 355  
 QY 181 CTGNGIMIANNGIERLRAGLIHDIEGIRHPPKPLGVDSKEVGEASIKVPOLKMEI 239  
 DB 356 CTGNGIMIANNGIERLRAGLIHDIEGIRHPPKPLGVDSKEVGEASIKVPOLKMEI 414

## RESULT 2

Q9H4B0 PRELIMINARY; PRT; 439 AA.  
 AC 09H4B0;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Putative stialoglycoprotease type 2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Uterus;  
 RA Chen J.M., Fortunato M., Barrett A.J.;  
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AJ295148; CAC14666.1; --  
 DR MEROPS; M22.004; --  
 DR Genew; HGNC:23075; OSGEP1.  
 DR GO; GO:0008450; F:O-stialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009050; Peptidase\_M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialyl.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Osialylc\_ptds; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TrifPfam; TIGR00329; gcp; 1.  
 KW Protease.  
 SO SEQUENCE 439 AA; 48040 MW; 44849372C784E41F CRC64;

Query Match 97.9%; Score 1213.5; DB 2; Length 439;  
 Best Local Similarity 90.2%; Pred. No. 1.3e-98;  
 Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

QY 1 LLAIVGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSMTSGGKAIIEHLAKQGNRF 60  
 DB 176 LLAIVGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSMTSGGKAIIEHLAKQGNRF 235  
 QY 61 HFDIKPPLHAKNCDPSFTGLOHVTDKIIMKKEKEGIEKGQILSSAADIAATVOHTMAC 98  
 DB 236 HFDIKPPLHAKNCDPSFTGLOHVTDKIIMKKEKEGIEKGQILSSAADIAATVOHTMAC 295  
 QY 99 ---EKGQILSSAADIAATVOHTMACHLVKTTRHAILFCCKORDLLPQNNAVLVASGVASN 155  
 DB 296 CRVEGQILSSAADIAATVOHTMACHLVKTTRHAILFCCKORDLLPQNNAVLVASGVASN 355  
 QY 156 FYIRALLETITNAQTCTLLCPPRLCTGNGIMIANNGIERLRAAGLIHDIEGIRHPPK 215

DB 356 FYIRALLETITNAQTCTLLCPPRLCTGNGIMIANNGIERLRAAGLIHDIEGIRHPPK 415  
 QY 216 PLGVDSKEVGEASIKVPOLKMEI 239  
 DB 416 PLGVDSKEVGEASIKVPOLKMEI 439

## RESULT 3

OSPEB4 PRELIMINARY; PRT; 414 AA.  
 AC 06PEB4;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=CZECH II; TISSUE=Mammary tumor;  
 RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RX Strausberg R.L., Pelngold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shmamen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo W.F., Casavant J.L., Schetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman W., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boufard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smalish D.E., Scherach A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=CZECH II; TISSUE=Mammary tumor;  
 RA Strausberg R.;  
 RL Submitted (SEP-2003) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; BC058172; AAH58172.1; --  
 DR GO; GO:0008450; F:O-stialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009050; Peptidase\_M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialyl.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Osialylc\_ptds; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TrifPfam; TIGR00329; gcp; 1.  
 KW Hypothetical protein.  
 SO SEQUENCE 414 AA; 44962 MW; 0418AD93C6FEB0CAB CRC64;

Query Match 87.6%; Score 1086; DB 2; Length 414;  
 Best Local Similarity 85.8%; Pred. No. 2.2e-87;  
 Matches 205; Conservative 15; Mismatches 19; Indels 0; Gaps 0;

QY 1 LLAIVGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSMTSGGKAIIEHLAKQGNRF 60  
 DB 176 LLAIVGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSMTSGGKAIIEHLAKQGNRF 235  
 QY 61 HFDIKPPLHAKNCDPSFTGLOHVTDKIIMKKEKEGIEKGQILSSAADIAATVOHTMAC 120

Dh 236 HFTINPWNKNCDFSTGLQHTIDKLTTHKEKEGIEKQILSSAADIAAVQHTAC 295  
Qy 121 HLKVTTRHAILFCCKQDLIPONNAVIVASGVASNFYIRALEILTNAQCTLLCPPEL 180  
Dh 296 HLAKEHTRHAILFCCKQNLSPANAVIVASGVASNFYIRKALEIVANATQCTLLCPPEL 355  
Qy 181 CTDNGIMIAMNGIERLRAGLIGLHDIIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239  
Dh 356 CTDNGIMIAMNGIERLRAGLIGLHDIIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 414  
RESULT 4  
08BLB6 PRELIMINARY; PRT; 414 AA.  
AC 08BLB6;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length  
DE enriched library, clone:B230219017 product:similar to PUTATIVE  
DE SIALOGLYCOPROTEIN TYPE 2.  
GN Name=Osgp11;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Meth. Enzymol. 303:19-44(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA RIKEN FANTOM Consortium;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RA The FANTOM Consortium;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Komuro H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Saeki N., Carninci P.,  
RA Komuro H., Akizawa J., Nishi K., Katsunai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,  
RA Yamamoto S., Matsunoto H., Sakaguchi S., Ikegami T., Kasahigaki K.,  
RA Fujikawa K., Inoue K., Togawa Y., Izawa M., Ohara B., Watanabe M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT cDNA sequencing pipeline with 384 multichannel sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [6]  
RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RA Aachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
RA Fukuchi S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
RA Horii F., Imotani K., Ishii Y., Itoh M., Kigawa I., Kusukawa T.,  
RA Kato H., Kawai J., Kojima Y., Kondo S., Komuro H., Kouda M., Koya S.,  
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohnato N., Okazaki Y.,  
RA Saito R., Saitoh K., Sakai C., Sakai K., Sakazume N., Sano H.,  
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,  
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK045669; BAC32450.1; -  
DR MEROPS; M22.004; -  
DR MGD; MGI:1919335; Osgp11.  
DR GO; GO:0008450; P:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR000905; Peptidase M22.  
DR InterPro; IPR009180; Pept M22 O-sialyl.  
DR Pfam; PF00814; Peptidase\_M22\_1.  
DR PRINTS; PRS004537; O-sialyl.  
DR PRINTS; PRS0789; OSIALOPTASE.  
DR ProDom; PD002367; Peptidase\_M22.  
DR TIGRPFAM; TIGR00329; gcp. 1.  
KM Protease.  
SQ SEQUENCE 414 AA; 44927 MW; 9A2D27B576C7DB24 CRC64;  
Query Match 87.0%; Score 1079; DB 2; Length 414;  
Best Local Similarity 85.8%; Pred. No. 9.1e-87;  
Matches 205; Conservative 14; Mismatches 20; Indels 0; Gaps 0;  
Qy 1 LALVGVSDFLILGKSLDIAPGMDLKVARRSLIKHPGCTMSGKAIEHLAKQGNRF 60  
Dh 176 LIALVGVSDFLILGKSLDIAPGMDLKVARRSLIKHPGCTMSGKAIEHLAKQGNRF 235  
Qy 61 HDIYKPLHAKNCDFSTGLQHTIDKLTTHKEKEGIEKQILSSAADIAAVQHTAC 120  
Dh 236 HFTINPWNKNCDFSTGLQHTIDKLTTHKEKEGIEKQILSSAADIAAVQHTAC 295  
Qy 121 HLKVTTRHAILFCCKQDLIPONNAVIVASGVASNFYIRALEILTNAQCTLLCPPEL 180  
Dh 296 HLAKEHTRHAILFCCKQNLSPANAVIVASGVASNFYIRKALEIVANATQCTLLCPPEL 355  
Qy 181 CTDNGIMIAMNGIERLRAGLIGLHDIIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239  
Dh 356 CTDNGIMIAMNGIERLRAGLIGLHDIIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 414  
RESULT 5  
Q9DON0 PRELIMINARY; PRT; 414 AA.  
AC Q9DON0;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Mus musculus 10 days embryo whole body cDNA, RIKEN full-length  
DE enriched library, clone:2610001M19 product:similar to PUTATIVE  
DE SIALOGLYCOPROTEIN TYPE 2.  
GN Name=Osgp11;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=whole body;  
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Meth. Enzymol. 303:19-44(1999).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RA the FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/9r.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes."  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/9r.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Wachihi M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer."  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
 RA Haraoka T., Hara A., Hayatsu N., Hiramoto K., Hirooka T., Hori F.,  
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,  
 RA Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,  
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Nunazaki R., Ohno M.,  
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,  
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN EMBL: AK011265; BAB27506.1; -  
 DR MEROPS: M22.004; -  
 DR MGD: MGI:191935; Osegep1.  
 DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO: GO:0008233; F:peptidase activity; IEA.  
 DR GO: GO:0008270; F:zinc ion binding; IEA.  
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR00905; Peptidase\_M22.  
 DR InterPro: IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam: PF00814; Peptidase\_M22\_1.  
 DR PIRSF: PIRSF004537; Osialglc\_ptds; 1.  
 DR PRINTS: PR00789; OSIALOPTASE.  
 DR ProDom: PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs: TIGR00329; gcp\_1.  
 DR TrEMBL: -  
 KW PROTEASE.  
 SQ SEQUENCE 414 AA; 44999 MW; 999BC689944DD24 CRC64;

QY 1 LIALVGVSDFLLLKSLDIAPGMDLKVARRSLIKHECSTMSGKAI EHLAKGNRP 60  
 DB 176 LIALVGVSDFLLLKSLDIAPGMDLKVARRSLIKHECSTMSGKAI EHLAKGNRP 235  
 QY 61 HFDIPEPLHAAKCPSPFGLQHTVDKIMKKEKRGIRKQILSSAADI AAVQHTMVC 120  
 DB 236 HFTINPQMNAKNCFFSPGLQHTVDKIMKKEKRGIRKQILSSAADI AAVQHTMVC 295  
 QY 121 HIKRTHRALFCRKORDLLPQNNAVLVASGVASNYIRALAEITLNATQCTLLCPPR 180  
 DB 296 HIAKTRHRLFLPCRKQNLSPNNAVLVASGVASNYIRKALEIVANAQCTLLCPPR 355  
 QY 181 CTDNGIMIMANGIERIRAGIIMHDI EGRYRKPCEPAGVDISKVGEASIKVPLQMEI 239  
 DB 356 CTDNGIMIMANGIERIRAGIIMHDI EGRYRKPCEPAGVDISKVGEASIKVPLQMEI 414  
 RESULT 6  
 O6AVN7 PRELIMINARY; PRT; 467 AA.  
 ID O6AVN7  
 AC 25-OCT-2004 (TrEMBLrel. 28, Created)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX PubMed=12477932; DOI=10.1073/pnas.242603999;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
 RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Umed T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Heitlen E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywiński M.I., Skalka U., Smalms D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Matra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Director MGC Project; to the EMBL/GenBank/DBJ databases.  
 RL EMBL: BC078974; AAH78974.1; -  
 DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO: GO:0008270; F:zinc ion binding; IEA.  
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR00905; Peptidase\_M22.  
 DR InterPro: IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam: PF00814; Peptidase\_M22\_1.  
 DR PIRSF: PIRSF004537; Osialglc\_ptds; 1.  
 DR PRINTS: PR00789; OSIALOPTASE.  
 DR ProDom: PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs: TIGR00329; gcp\_1.  
 DR PROSITE: PS00435; PEROXIDASE\_1; UNKNOWN\_1.  
 KW Hypothetical protein.



FT NON TER 467 467  
 SQ SEQUENCE 467 AA; 50799 MW; 474E1B1959B9AC0 CRC64;  
 Query Match 79.3%; Score 983; DB 2; Length 467;  
 Best Local Similarity 87.9%; Pred. No. 3.1e-78;  
 Matches 188; Conservative 9; Mismatches 17; Indels 0; Gaps 0;

QY 1 LLAIVGVSDPFLLGKSLDIAPGMDLVKARRLSLIKHEPCSTMSGKAI EHLAKGNRF 60  
 DB 176 LLAIVGVSDPFLLGKSLDIAPGMDLVKARRLSLIKHEPCSTMSGKAI EHLAKGNRF 235  
 QY 61 HFDIKPPLHAKNCDFPTGLQHTVDKLIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 120  
 DB 236 HFTINPPEQNAKNCDFPTGLQHTVDKLIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 295  
 QY 121 HLKVRTRRAILFCCKQDLPQNNNAVVASGVASNFYIRRALEILTNAQCTLLCPPRL 180  
 DB 296 HLKVRTRRAILFCCKQDLPQNNNAVVASGVASNFYIRRALEILTNAQCTLLCPPRL 355  
 QY 181 CTDNGIMIANWGIERLRAGLGIHDIIEGIRYEPK 214  
 DB 356 CTDNGIMIANWGIERLRAGLGIHDIIEGIRYEPK 389

RESULT 7  
 Q96NH5 PRELIMINARY; PRT; 364 AA.  
 ID Q96NH5  
 AC Q96NH5;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Hypothetical protein FLJ30879.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX PubMed=14702039; DOI=10.1038/ng1285;  
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura T., Makita H.,  
 Sekine M., Ohyashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
 Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,  
 Murakami K., Yasuda T., Iwanaga T., Magatsuna M., Shimizu A.,  
 Sudo H., Hosoiri T., Kaku Y., Kodaera H., Kondo H., Sugawara M.,  
 Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
 Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,  
 Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,  
 Tanai H., Kimata M., Matsumae M., Hirakawa S., Chiba Y., Ishida S.,  
 Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,  
 Kanehori K., Takahashi-Fuji A., Hara H., Tanase T., Nomura Y.,  
 Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
 Nishiyama K., Matsunawa H., Ichihara T., Shiohara T., Sano S.,  
 Moriya S., Momiya H., Satoh N., Takami S., Teraishi N., Sano S.,  
 Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
 Hishiyaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
 Yamashita Y., Komiyama M., Tashiro H., Tanigami A., Fujimori T.,  
 Oono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Omori Y.,  
 Okabe A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
 Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
 Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,  
 Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,  
 \*Complete sequencing and characterization of 21,243 full-length human  
 cDNAs.\*  
 RT Nat. Genet. 36:40-45 (2004).  
 RL EMBL; AK055441; BAB70923.1; -.  
 DR MEROPS; M22.004; -.

DR GO; GO:0008450; F10-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F1-peptidase activity; IEA.  
 DR GO; GO:0008270; F2-zinc ion binding; IEA.  
 DR GO; GO:0006508; P1-proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000905; Peptidase M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF04537; Osialglc\_pptds; 1.  
 DR PRINTS; PR00789; OSIALOPRASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 KW Protease.  
 SQ SEQUENCE 364 AA; 39528 MW; E0E605A07D0EC3D6 CRC64;  
 Query Match 77.4%; Score 960; DB 2; Length 364;  
 Best Local Similarity 98.4%; Pred. No. 2.5e-76;  
 Matches 186; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 LLAIVGVSDPFLLGKSLDIAPGMDLVKARRLSLIKHEPCSTMSGKAI EHLAKGNRF 60  
 DB 176 LLAIVGVSDPFLLGKSLDIAPGMDLVKARRLSLIKHEPCSTMSGKAI EHLAKGNRF 235  
 QY 61 HFDIKPPLHAKNCDFPTGLQHTVDKLIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 120  
 DB 236 HFDIKPPLHAKNCDFPTGLQHTVDKLIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 295  
 QY 121 HLKVRTRRAILFCCKQDLPQNNNAVVASGVASNFYIRRALEILTNAQCTLLCPPRL 180  
 DB 296 HLKVRTRRAILFCCKQDLPQNNNAVVASGVASNFYIRRALEILTNAQCTLLCPPRL 355  
 QY 181 CTDNGIMIA 189  
 DB 356 CTDNGIMIA 364

RESULT 8  
 Q8JFW3 PRELIMINARY; PRT; 404 AA.  
 ID Q8JFW3  
 AC Q8JFW3;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE SI:d2211013.4 (Novel glycoprotease).  
 GN Name=d2211013.4;  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 NC NCB1\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Babage A.;  
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL591593; CND43471.1; -.  
 DR MEROPS; M22.004; -.  
 DR GO; GO:0008450; F10-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F1-peptidase activity; IEA.  
 DR GO; GO:0008270; F2-zinc ion binding; IEA.  
 DR GO; GO:0006508; P1-proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000905; Peptidase M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF04537; Osialglc\_pptds; 1.  
 DR PRINTS; PR00789; OSIALOPRASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 KW Protease.  
 SQ SEQUENCE 404 AA; 43956 MW; 3A6B1DD050737B35 CRC64;  
 Query Match 60.6%; Score 751; DB 2; Length 404;  
 Best Local Similarity 61.6%; Pred. No. 7.9e-58;  
 Matches 146; Conservative 32; Mismatches 59; Indels 0; Gaps 0;

QY 1 LLAIVOGSDFLFLLKSLID IAPGDMADKYARSLIKHPSCSMWGSAI EHLAKONRF 60  
 :  
 Db LLALAKGIDEHLLDQTLDEAGDTIDTKARRSLRNHPCCGLTSGOAIERLAKSBDRL 22  
 164  
 QY 61 HFDIKPRLHAKNCDPSFTGLQHVTDKIIMKEKEBEGIEKGQILSSAADIATVQHTMAC 120  
 :  
 Db AEFHISFGQNVDCNFSFAGLRQTQLGALINKKEBGVEAGCLSTCVKDIAASQHTVAS 283  
 224  
 QY 121 HLVRKTHALLFCRQRDLFPNNNAVVGAGVANFYIRALELTMAOTCTLLCPPRLL 180  
 :  
 Db HLAKRTHALLFCCKRGKLPEONPRTLIVSGVANSYIROLITITDATSEVALHLLCPSPKF 343  
 284  
 QY 181 CTDNGIMTAMNGIERBLRAGLGIHLHDIGRIYEPKPCPLGVNISKEVGEASIKVPOLKM 237  
 :  
 Db CTDNGIMTAMNGIERBLRQCKGILLSYEEAVSYEKAPLGIDLTSVEKENALKVPERKLK 400  
 344

RESULT 9			
ID	08JFR7		
AC	08JFR7	PRELIMINARY;	PRT; 404 AA.
DT	01-OCT-2002 (TREMBLrel. 22, Created)		
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)		
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)		
DE	SI:d72B14.6 (Novel glycoprotease).		
GN	Name=SI:d72B14.6;		
OS	Brachydanio rerio (Zebrafish) (Danio rerio).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes		
OC	Cyprinidae; Danio.		
OX	NCBI_TaxID=7955;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RA	Clark G.;		
RL	Submitted (JUN-2003) to the EMBL/GenBank/DBD databases.		
DR	EMBL; AL672217; CAD3443.1; ..		
DR	MEOPSP; M22.004; ..		
DR	GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.		
DR	GO:0008233; F:peptidase activity; IEA.		
DR	GO:0008270; F:zinc ion binding; IEA.		
DR	GO:0006508; P:proteolysis and peptidolysis; IEA.		
DR	InterPro; IPR009805; Peptidase_M22.		
DR	InterPro; IPR009180; Pept_M22_Ostia1g1.		
DR	Pfam; PF00814; Peptidase_M22_1.		
DR	PIRSF; PIRSF004537; Ostia1gic_ptds; 1.		
DR	PRINTS; PR00789; OSIALOPTAS.		
DR	Prodom; PD002367; Peptidase_M22_1.		
DR	TIGRFAMS; TIGR00329; gcp_1.		
SC	SEQUENCE. 404 AA; 44027 MW; 6F98653A651860F CRC64;		

	Query Match	60.2%;	Score 747;	DB 2;	Length 404;		
	Best Local Similarity	61.2%;	Pred. No. 1.8e-57;				
	Matches	145;	Conservative	33;	Mismatches	59;	
				Indels	0;	Gaps	0
Oy	1	LIALVQGVSDPILLAGKSLDIAPGDMLDKYARRLSLIKHPECSYTMGGGAIEHLAQGNRF	60				
Db	164	LIALLAKGIDFELLGOTLDEAAGDPTLDIKIARLSLRNHPCCITLSSGQAIEHLAREGDL	223				
Oy	61	HFDIKPLFIHAAKCDPSFTGLQHVWTDKIIMKEKEEGEKGQILSSADIAATVOHTMAC	120				
Db	224	AFFHISPMGQVYDCNCFSPAGLRTOQTGAINKKEKEGEVAQGPFICVDIIAASQHTVAS	283				
Oy	121	HLVKRTRHALFECKORDLLPONNAVLVASGVASNFIYIRALEELTNATQCTLLCPPEPL	180				
Db	284	HLAKRTHRALIFCCKSGKLPLEQNPTLIVSGAVASNEYIRQLIKITDTATGHIHLCPSPKF	343				
Oy	181	CTDNGCIMANNGIIEFLRACGLIHIDIEGRVSKPKPLGDISEKVEASIKYPOIXM	237				
Db	344	CTDNCVMIANNGIIEFLRKQGLISYEESVYPKPALPDIGTISEVEAAIKVPKUL	400				

ID	070918		
AC	070918	PRELIMINARY;	PRT; 401 AA.
DT	01-MAR-2004	(TRENBLrel. 26, Created)	
DT	01-MAR-2004	(TRENBLrel. 26, Last sequence update)	
DE	01-MAR-2004	(TRENBLrel. 26, Last annotation update)	
DR	AGCP14990	(Fragment).	
GN	Name=agCG46164; ORFNames=ENSNACG000000007922;		
OS	Anopheles gambiae str. PEST.		
OC	Bakayocra; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
NC	Nesoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anophelae		
OX	NCEI_Taxid=180454;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=PEST;		
RA	Anopheles Genome Sequencing Consortium;		
RL	Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.		
CC	-!- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is preliminary data.		
CC	EMBL; AAB01008900; EAA09387.1; -		
DR	GO; GO:0008450; P/O-stialoglycoprotein endopeptidase activity; IEA.		
DR	GO; GO:0008270; Zinc ion binding; IEA.		
DR	GO; GO:0005508; Proteolysis and peptidolysis; IEA.		
DR	InterPro; IPR000905; Peptidase M22.		
DR	InterPro; IPR009180; Pept_M22_OstAlgl.		
DR	Pfam; PF00814; Peptidase M22; 1.		
DR	PIRSF; PIRSF004537; Ostialgic_ptctds; 1.		
DR	PRINTS; PR00789; OSIALOCTPASE.		
DR	Prodom; PD002367; Peptidase_M22; 1.		
DR	TIGRfams; TIGRF00329; gcp; 1.		
DR	NON_TBR		
SQ	SEQUENCE	1	1
401	AA; 43849 MW; CC942672D3D4E1 CRC64;		

	Query Match	29.2%	Score 361.5;	DB 2;	Length 401,
	Best Local Similarity	35.3%;	Pred. No. 1.9e-23;		
	Matches 90;	Conservative 44;	Mismatches 103;	Indels 11;	Gaps 4
Oy	1 LIALVQGVDFLLKGSKDLDIAPGDMIDKVARRSLTIKKPECSYWSGAIEHTIAKQG--	57			
Dd	147 LIIVVESTARRRLGETLDDPAGELDKIARLRKLKNAYKAQNMSGGALTEAAQAQAKX	206			
Oy	58 --NRHFPIKDIPRLHRAKNKDSFTGLQHVTDKIIIMKKKEEGEIKGQLSSAADIATVQ	115			
Dd	207 DTSAVEPPL--FLSKYRRCQGFSPFAGLKNTARRHLIERSTLIHADALLPDYEAFCAERL	264			
Oy	116 HTMAACHLVKRTTHRAILFCQRDL--PONNAVLVASGVASNFYRRRLBITTNAQTCL	173			
Dd	265 KGVTRHMLHFGRAILEYCERRKLTFSDAEPHRSLVYSGVACADVIFNALSSMAAQFGYST	324			
Oy	174 LCPPRLCTDNGMIAMNIGIRLTA--GLGIHLDEGRIRYFKCPGLVDISMEVEASLIK	231			
Dd	325 YRPPEKKLTNDGTWIAWNMGKELFAKDPTAEITTKYEOVDISGCKPIGSLIIDVXEANITA	384			
Oy	232 VPOLKMEI 239				
Dd	385 CKMAKVDI 392				

```

RESULT 11
Q960S6
ID Q960S6 PRELIMINARY; PRT; 323 AA.
AC Q960S6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE LD37221P.
GN ORFNames=CG14231;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;

```



DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Osialglc\_ptds; 1.  
 DR PRINTS; PR00789; OSIALOPTASE  
 DR PRODOM; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 DR SEQUENCE 409 AA; 45328 MW; 979766E7D155538 CRC64;  
 SQ  
 Query Match 26.8%; Score 332.5; DB 2; Length 409;  
 Best Local Similarity 35.9%; Pred. No. 6.9e-21;  
 Matches 83; Conservative 38; Mismatches 97; Indels 13; Gaps 5;  
 QY 2 LATVQGVSDFLILGKSLDIAPGMDLVKVARSLIKHECSTMGCKAIEHLAK-QGNRF 60  
 DB 167 LVVANGGRLTLTGQTDADAGEAFDKIGRLRLHLPEYLLMNGRAIEHNAQASDPL 226  
 QY 61 HFDIKPPLHAKNCDFSTGLOHTDKIMKEKEGEGKQIILSSADIAATVQHTMAC 120  
 DB 227 AYEPLPLAQORNCNCFSPAGIKNNSFRAIRERERPPDGVISNYGDFCGLRSVS 286  
 QY 121 HLYKTRRAILFC--KORDLLPNNNAVIVASGVASNYIRRALEILTNATQCTLLCP 178  
 DB 287 HLMHRTPRALYCLLPHKQLFGDTPPTLVMSGVANNDAIYANILHAAQICRSFRPSK 346  
 QY 179 RLCTDNGIMIAMNGIERLRAGILHDIE-GIRYEPKCPGLGVDSKEVGEA 228  
 DB 347 RYCSDNGVMIMHVEQL-----LQDKKASTRYDYD---SIDIGSGNFA 388  
 RESULT 13  
 022145 PRELIMINARY; PRT; 480 AA.  
 AC 022145; OSVWL2;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 05-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Putative O-sialoglycoprotein endopeptidase (Sialoglycoprotease GCP1).  
 GN Name:At2g45270; Synonyms:GCP1;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,  
 RA Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R.,  
 RA Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RA SEQUENCE FROM N.A.  
 RA Town C.D., Kaul S.;  
 RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RA SEQUENCE FROM N.A.  
 RA Haussehl K., Adamska I.;  
 RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RA SEQUENCE FROM N.A.  
 RA Yamada K., Ito S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
 RA Seki M., Shim P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.  
 RN [5]  
 RA SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,  
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,

RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
 RA Seki M., Shim P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RL Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AC002387; AAB82636.2; -  
 DR EMBL; AY024338; AAK00530.1; -  
 DR EMBL; AY063864; AAL36220.1; -  
 DR EMBL; AY117283; AAM51358.1; -  
 DR PIR; E84888; E84888.  
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Osialglc\_ptds; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR PRODOM; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 DR PROTEASE.  
 KW  
 SQ SEQUENCE 480 AA; 52995 MW; 20DD6A86ACC1FFAD CRC64;  
 Query Match 26.1%; Score 324; DB 2; Length 480;  
 Best Local Similarity 34.1%; Pred. No. 4.7e-20;  
 Matches 86; Conservative 34; Mismatches 86; Indels 46; Gaps 6;  
 QY 1 LIALVQGVSDFLILGKSLDIAPGMDLVKVARSLIKHECSTMGCKAIEHLAKGNRF 60  
 DB 224 LILVLAHKLGOYQLGTGVDDAIGEAAPDKTAKVGLDMH-----RSGGPVABEILLEGDAK 278  
 QY 61 HFDIKPPLHAKNCDFSTGLOHTDKIMKEKEGEGKQIILSSA-----ADIVA 112  
 DB 279 SVKENVPMKYHDCNFSYAGLQYRLALEAKE---IDAKCPVSSATNEDRRNRADYIA 334  
 QY 113 TVQHTMACILVARTRAILFC--KORDLLPNNNAVIVASGVASNYIRRALEILTNATQCT 172  
 DB 335 SGRVAVLHLBERKCEAIDMALE---LEPSIKHNVISGVASNYIRRLNNTVENKGNL 391  
 QY 173 LILCPPELCTDNGIMIAMNGIERLRAGILHDIEGRYE-----PKCP 216  
 DB 392 LVCPPPLCTDNGVMIVAGTGLEHFRVG-----RYDPPPEATEPEDYVYDLBRWP 441  
 QY 217 LGVDISKVEGEA 228  
 DB 442 LGBEVAKGRSEA 453  
 RESULT 14  
 0677H2 PRELIMINARY; PRT; 255 AA.  
 AC 0677H2;  
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE O-sialoglycoprotein endopeptidase (Fragment).  
 GN Hyacinthus orientalis.  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Hyacinthaceae;  
 OC Hyacinthus.  
 OC NCBI\_TaxID=82025;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RA TISSUE=Floral meristem 5-10 days when regenerated in vitro;  
 RA Pan J.H., Ma Y., Zhang X.S.;  
 RT mRNA, expressed during regeneration of floral bud.  
 RL Submitted (AUG-2003) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AY389554; AAT08652.1; -  
 DR InterPro; IPR009095; Peptidase\_M22.  
 DR PRINTS; PR00789; OSIALOPTASE.

DR Prodom: PD002367; Peptidase\_M22; 1.  
FT NON\_TER 1 1  
FT NON\_TER 255 255  
SQ SEQUENCE 255 AA; 28516 MW; 529D2266C284106 CRC64;

Query Match 24.2%; Score 299.5; DB 2; Length 255;  
Best Local Similarity 37.4%; Pred.No. 3.3e-18;  
Matches 79; Conservative 26; Mismatches 79; Indels 27; Gaps 5;

QY 14 LGSGLDIAPGDMLDKVARRLSLIKHPECSMSGKAIIEHLAKQNRPHFDIPPLHAKN 73  
DB 1 LGTTIDDAIGAYDKTARWLG- - - - -DMGKGGGALIEELAEGDANSYKFIIPKQKQ 55

QY 74 CDFSEFTGLQHTVDKIIIMKEKEEGIEKQILSSA- - - - -ADIAATVQHTNACHLVKRT 126  
DB 56 CNFSYVAGLK- - - - -TVRLAIESRNICIDKSPISLATCHDRMRADIASFQAVLHLEDR 113

QY 127 HRAILFCQKQRLDLPQNNAVLVASGVASNFYIRALIELTNATQCTLCPPLCTDNGI 186  
DB 114 ARAIEMALEIEPTVO- - - - -CLVVGSGVASNKYVRSRLNHLVKTGLVCPPLCTDNGV 170

QY 187 MIANGIERLARGILHDIGIREYRCPPL 217  
DB 171 MVAWTGIENFLVG- - - - -RYDPPPPV 191

## RESULT 15

092LH8

PRELIMINARY; PRT; 360 AA.

AC 092LH8; 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE 01-MAR-2004 (TREMBLrel. 26, Last annotation update)

DE PROBABLE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57).

GN ORFNames=SMC03230;

OS Rhizobium meliloti (Sinorhizobium meliloti);

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.

OX NCBI\_TaxID=382;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=1021;

RX MEDLINE=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;

RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Barut J.,

RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,

RA Godrie T., Goffeau A., Kahn D., Kiss B., Lelaure V., Masny D.,

RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,

RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;

RT "Analysis of the chromosome sequence of the legume symbiont

RT Sinorhizobium meliloti strain 1021."

RT Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).

DR EMBL; AL591792; CAC47657.1; -.

DR GO; GO:0016787; F:hydrolase activity; IEA.

DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.

DR GO; GO:0008270; F:zinc ion binding; IEA.

DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR InterPro; IPR000905; Peptidase\_M22.

DR InterPro; IPR001801; Pept\_M22\_Osialgl.

DR Pfam; PF00814; Peptidase\_M22; 1.

DR PIRSF; PIRSF004537; Osialglc\_ptide; 1.

DR PRINTS; PR00789; OSIALOPTASE.

DR Prodom; PD002367; Peptidase\_M22; 1.

DR TIGRPFMS; TIGR00329; gcp; 1.

DR TIGRPFMS; TIGR00329; gcp; 1.

KW Complete proteome; Hydrolase.

SQ SEQUENCE 360 AA; 37906 MW; A07F946AB562EA86 CRC64;

Query Match 21.2%; Score 262.5; DB 2; Length 360;

Best Local Similarity 32.8%; Pred.No. 9e-15; DB 2; Length 360;

Matches 78; Conservative 32; Mismatches 85; Indels 43; Gaps 7;

QY 2 LALVQGVDFLLKSLDIAPGDMLDKVARRLSLIKHPECSMSGKAIIEHLAKQNRPH 61  
DB 145 LILVKGVEYRMKGTITDDAIGEAADFRTAKLGL-PYP- - - - -GGPAAVRAQNGNMR 197

QY 62 FDIKPPHHAKNCDFSEFTGLQ- - - - -HYTDKIIMKEKEEGIEKQILSSAADI 110  
DB 198 FDFPRPLVGDAFLDFSGIKTAVRQAQSLGPTDQDI- - - - -ADV 239

QY 111 AATVQHTMACHLVKRTHRAILFCQKQRLDLPQNNAVLVASGVASNFYIRALIELTNATQ 170  
DB 240 CASFORAISRTRLDRVGRGLKRR-ADFASVDPALVAVAGVANQTLRTTQLSCDEHG 298

QY 171 CTLCPPLCTDNGIMIAMNGIERLARGILHDIGIREYRKC- - - - -PLGVDISKV 226  
DB 299 FRTIAPPLQCTDNNAMIMAGARLAAGL- - - - -PADGLDAAPRRRWPDLSEAKALIG 352

Search completed: November 10, 2005, 16:52:03  
Job time : 58.9315 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 15:51:52 ; Search time 19.2239 Seconds  
(without alignments)  
928.069 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 LIALVQGVSDPFLIGKSLDI.....DISKVEGASIKVPOLKMEI 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1240	100.0	267	US-10-067-443-22	Sequence 22, Appl
2	1240	100.0	414	US-10-067-443-2	Sequence 2, Appl
3	1213.5	97.9	439	US-10-067-443-19	Sequence 19, Appl
4	326	26.3	463	US-10-067-443-3	Sequence 3, Appl
5	247	19.9	350	US-09-540-236-2726	Sequence 2726, Ap
6	231	18.6	401	US-09-252-991A-17372	Sequence 17372, A
7	212	17.1	357	US-09-543-681A-6513	Sequence 6513, Ap
8	209.5	16.9	342	US-08-087-797-3	Sequence 3, Appl
9	205	16.5	343	US-09-489-039A-9221	Sequence 9221, Ap
10	201.5	16.2	421	US-10-067-443-4	Sequence 4, Appl
11	201.5	16.2	421	US-10-067-443-28	Sequence 28, Appl
12	200	16.1	325	US-08-087-797-2	Sequence 2, Appl
13	197	15.9	363	US-09-107-532A-6609	Sequence 6609, Ap
14	193.5	15.6	327	US-10-067-443-5	Sequence 5, Appl
15	190	15.3	336	US-08-987-121A-4	Sequence 4, Appl
16	189	15.2	374	US-09-710-279-728	Sequence 728, App
17	187	15.1	344	US-09-602-777A-148	Sequence 148, App
18	186	15.0	335	US-08-961-083-52	Sequence 52, Appl
19	186	15.0	335	US-09-536-784-52	Sequence 5, Appl
20	186	15.0	368	US-09-134-001C-3909	Sequence 3909, Ap
21	184	14.8	336	US-09-066-512-2	Sequence 2, Appl
22	182	14.7	336	US-09-583-110-4857	Sequence 4857, Ap
23	181.5	14.6	143	US-09-328-352-4387	Sequence 4387, Ap
24	177	14.3	336	US-07-107-453-4221	Sequence 4221, Ap
25	174.5	14.1	344	US-09-198-452A-213	Sequence 213, App
26	174.5	14.1	360	US-09-438-185A-196	Sequence 196, App
27	174	14.0	341	US-09-149-624-2	Sequence 2, Appl

28	173	14.0	366	US-09-134-000C-4956	Sequence 4956, Ap
29	148.5	12.0	340	US-10-067-443-6	Sequence 6, Appl
30	117	9.4	292	US-09-724-623-81	Sequence 81, Appl
31	94.5	7.6	480	US-09-583-110-5050	Sequence 5050, Ap
32	94.5	7.6	481	US-09-107-433-3197	Sequence 3197, Ap
33	89.5	7.2	214	US-09-328-352-4609	Sequence 4609, Ap
34	83	6.7	42	US-09-302-540-13841	Sequence 13841, A
35	81.5	6.6	1087	US-08-264-002-5	Sequence 5, Appl
36	81	6.5	1072	US-09-902-540-15572	Sequence 15572, A
37	79	6.4	445	US-08-083-945C-2	Sequence 2, Appl
38	79	6.4	445	US-08-083-945C-7	Sequence 7, Appl
39	79	6.4	578	US-09-252-991A-3318	Sequence 3318, A
40	79	6.4	1137	US-09-538-092-968	Sequence 968, App
41	78	6.3	334	US-08-530-950-4	Sequence 4, Appl
42	78	6.3	334	US-08-576-240-2	Sequence 2, Appl
43	78	6.3	334	US-08-888-429A-4	Sequence 4, Appl
44	78	6.3	334	US-09-149-879-4	Sequence 4, Appl
45	78	6.3	334	US-09-057-009-4	Sequence 4, Appl

ALIGNMENTS

```
RESULT 1
US-10-067-443-22
; Sequence 22, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYPEPTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266, 518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282, 814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 22
; LENGTH: 267
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-443-22

Query Match 100.0%; Score 1240; DB 4; Length 267;
Best Local Similarity 100.0%; Pred. No. 5.6e-145;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIALVQGVSDPFLIGKSLDIAPGDMIDKVARRLSLIKPECGTMSGGKAIIEHLAKGNRF 60
|||
DB 29 LIALVQGVSDPFLIGKSLDIAPGDMIDKVARRLSLIKPECGTMSGGKAIIEHLAKGNRF 88
|||
QY 61 HFDIKPELHAKNCDFSPFGLOHVTDKITMKKEKEGIRKGOILSSAADIATVQHTMAC 120
|||
DB 89 HFDIKPELHAKNCDFSPFGLOHVTDKITMKKEKEGIRKGOILSSAADIATVQHTMAC 148
|||
QY 121 HLVKETTRAILFCCKORDLLPQNNAVLVASGVASNYFIRRALEILTNAQTCTLLCPPRL 180
|||
DB 149 HLVKETTRAILFCCKORDLLPQNNAVLVASGVASNYFIRRALEILTNAQTCTLLCPPRL 208
|||
QY 181 CTDNQIMTAMNCEIEHLRAGLGILHDIIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 239
|||
DB 209 CTDNQIMTAMNCEIEHLRAGLGILHDIIEGIRYEPKCPGLVDISKVEGASIKVPOLKMEI 267
|||

RESULT 2
US-10-067-443-2
; Sequence 2, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
```





SEQ ID NO 2726  
LENGTH: 350  
TYPE: PRT  
ORGANISM: M.catarrhalis  
US-09-540-236-2726

Query Match 19.9%; Score 247; DB 4; Length 350;  
Best Local Similarity 31.5%; Pred. No. 1,1e-21;  
Matches 70; Conservative 34; Mismatches 80; Indels 38; Gaps 6;

QY 1 LALVQGVSDFLILGKSLDIAPGDMLDKVARRLSLIKHPECSMTSGGKAIENHAKQGNRF 60  
DB 149 MLVRADSGVQIILGESSIDDAVGECPDXTAKMLK-PYF-----GGPIETKLKNGNH 201  
QY 61 HFIDKPLHAKNCDPFTGLQHTVDKTIIMKKEKEGIEKGQILSSADIATVQHTMAC 120  
DB 202 AYELPRPWOH-KGLDFSPSKMTAIHNLIKOTPNQSDP-----ATRADIAASFVAAYD 255  
QY 121 HLVRKTHRAILFCRKORDLPQNNAVLVASGVASNFYIRALEILTNAQCTLLCPPL 180  
DB 256 TLVKKCTVALQMTGIRQ-----LVVAGVSAQTLEKRLTETLRQIDASVYVAPTEL 307  
QY 181 CTDNGIMIAMNGIERLAG-----GLIHD 205  
DB 308 CTDNGAMIAVAGFCRLSRQSDDLAVRCIPRMDMTMLGIHYD 349

## RESULT 6

US-09-252-991A-17372  
Sequence 17372, Application US/09252991A

PATENT NO. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 17372  
LENGTH: 401  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17372

Query Match 18.6%; Score 211; DB 4; Length 401;  
Best Local Similarity 33.8%; Pred. No. 1.3e-19;  
Matches 73; Conservative 34; Mismatches 87; Indels 22; Gaps 8;

QY 2 LALVQGVSDFLILGKSLDIAPGDMLDKVARRLSLIKHPECSMTSGGKAIENHAKQGNRF 61  
DB 202 LVRVDGIGRYGLTESYDDAAGEAFDXTAKLIGL-GYP-----GGPEIARLAEGTPER 254  
QY 62 FDIKPLHAKNCDPFTGLQHTVDKTIIMKKEKEGIEKGQILSSADIATVQHTMAC 121  
DB 255 FVPRPMTDRPGDLFPFSGLTFTPLN-TWQRCVEAGDSEQ---TRCDIALAFQAVAVET 310  
QY 122 LVRKTHRAILFCRKORDLPQNNAVLVASGVASNFYIRALEILTNAQCTLLCPPL 181  
DB 311 LLIKCRRL--KOTGL--LVLAGVSAQNALSGLEKMGKGVFVARPRPC 362  
QY 182 TDNGIMIAMNGIERLAGIILHIDIEGIRYEPKPL 217  
DB 363 TDNGAMIAVAGCQRLLAG---QHGPRAISVQPRWF 395

## RESULT 7

US-09-543-681A-6513  
Sequence 6513, Application US/09543681A

PATENT NO. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
FILE REFERENCE: 2709.1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 6513  
LENGTH: 357  
TYPE: PRT  
ORGANISM: Proteus mirabilis  
US-09-543-681A-6513

Query Match 17.1%; Score 212; DB 4; Length 357;  
Best Local Similarity 30.3%; Pred. No. 2.4e-17;  
Matches 66; Conservative 32; Mismatches 90; Indels 30; Gaps 7;

QY 2 LALVQGVSDFLILGKSLDIAPGDMLDKVARRLSLIKHPECSMTSGGKAIENHAKQGNRF 61  
DB 159 LLSVTGIGRYTLGESSIDDAAGEAFDXTAKLIGL-DYP-----GGVLSKMAQGVGR 211  
QY 62 FDIKPLHAKNCDPFTGLQHTVDKTIIMKKEKEGIEKGQILSSADIATVQHTMAC 121  
DB 212 FVPRPMTDRPGDLFPFSGLTFTPLN-TWQRCVEAGDSEQ---TRCDIALAFQAVAVET 263  
QY 122 LVRKTHRAILFCRKORDLPQNNAVLVASGVASNFYIRALEILTNAQCTLLCPPL 180  
DB 264 LLIKCRRA-----LEQGFRLVMAQVSAQNTLRKAMMIMEQGVFVARPRPC 314  
QY 181 CTDNGIMIAMNGIERLAGIILHIDIEGIRYEPKPL 217  
DB 315 CTDNGAMIAVAGCQRLLAGIILHIDIEGIRYEPKPL 348

## RESULT 8

US-08-087-797-3  
Sequence 3, Application US/08087797

PATENT NO. 5543312  
GENERAL INFORMATION:  
APPLICANT: Mellors, Alan  
APPLICANT: Io, Reggie Y.C.  
TITLE OF INVENTION: Pasteurella Haemolytica  
TITLE OF INVENTION: Glycoprotease  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Bell, Seltzer, Park & Gibson, P.A.  
STREET: 1211 East Morehead Street,  
CITY: Charlotte  
STATE: No. 5543312  
COUNTRY: United States  
ZIP: 28234  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087,797  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Layton, Jr., Samuel G.  
REGISTRATION NUMBER: 22807  
REFERENCE/DOCKET NUMBER: 3374-80  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 704 377 1561  
TELEFAX: 704 334 2014

; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 342 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-087-797-3

Query Match 16.9%; Score 209.5; DB 1; Length 342;  
Best Local Similarity 30.9%; Pred. No. 4.7e-17;  
Matches 67; Conservative 30; Mismatches 97; Indels 23; Gaps 7;

QY 2 LALVGVSDPFLGSLDIPGDMDKVARRSLIKHEGCTMSGGKAIEHLAKOGRH 61  
DB 142 LITVIGIQVEYELGESIDDAAGEADPKTKLGL-DYP-----GGPMLSKMASQCTBGR 194  
QY 62 FDIKPEPLHAAKNCDFSFTGLQHTVDKIIIMKEKEGIEKGQILSSAADIATVQHTMACH 121  
DB 195 FVFPFRPMTDRPGLDPSFGSLKTFPAANTIRDNXXXXGXTHDDQ---TRADIARAFEDAVVDT 251  
QY 122 LVKTRTHALIFCKQDILLPÖNN-AVLVAGGVASNFYIRRAELITNATQCTLLCPPEPL 180  
DB 252 LMICKKRA-----LDQTFKRLVMAAGVSAANTIRAKLAEEMKKRKGVEVFYARPEF 302  
QY 181 CTDNGIMAMNGIERLRAGLGIHDIEGIRYEPKCP 217  
DB 303 CTDNGAMTAYAGVWRVFKA--GATADL-GVSVRPRMPL 336

## RESULT 9

US-09-489-039A-9221  
; Sequence 9221, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 9221  
; LENGTH: 343  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
; US-09-489-039A-9221

Query Match 16.5%; Score 205; DB 4; Length 343;  
Best Local Similarity 30.3%; Pred. No. 1.7e-16;  
Matches 66; Conservative 29; Mismatches 93; Indels 30; Gaps 7;

QY 2 LALVGVSDPFLGSLDIPGDMDKVARRSLIKHEGCTMSGGKAIEHLAKOGRH 61  
DB 148 LITVIGIQVEYELGESIDDAAGEADPKTKLGL-DYP-----GGPMLSKMASQCTBGR 200  
QY 62 FDIKPEPLHAAKNCDFSFTGLQHTVDKIIIMKEKEGIEKGQILSSAADIATVQHTMACH 121  
DB 201 FVFPFRPMTDRPGLDPSFGSLKTFPAANTIRSDNDE-----QTRADIARAFEDAVVDT 252  
QY 122 LVKTRTHALIFCKQDILLPÖNN-AVLVAGGVASNFYIRRAELITNATQCTLLCPPEPL 180  
DB 253 LMICKKRA-----LEQTFKRLVMAAGVSAANTIRAKLAEEMKKRKGVEVFYARPEF 303  
QY 181 CTDNGIMAMNGIERLRAGLGIHDIE-GIRYEPKCP 217  
DB 304 CTDNGAMTAYAGVWRVFKA--KAEIGVTVRPRMPL 337

RESULT 10  
US-10-067-443-4

; Sequence 4, Application US/10067443  
; Patent No. 6642041  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I  
; TITLE OF INVENTION: SPINAL CORD, MP-1  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 421  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; US-10-067-443-4

Query Match 16.2%; Score 201.5; DB 4; Length 421;  
Best Local Similarity 27.5%; Pred. No. 6.3e-16;  
Matches 69; Conservative 48; Mismatches 99; Indels 35; Gaps 7;

QY 1 LALVGVSDPFLGSLDIPGDMDKVARRSLIKHEGCTMSGGKAIEHLAKOGRH 58  
DB 163 LISVADVEKFTLQGVSGSRECEIDKVARQGLD-GSEFSDIHGAIVELASRASA 221  
QY 59 -RHFIDIKPEPLHAAKNCDFSFTGLQHTVDKIIIMKEKEGIEKGQILSSAADIATVQHT 117  
DB 222 GHLRVPILFENVPKANMFPQIKGSYLNILRLRKNSETSID-----IPDFCASLQNT 274  
QY 118 MACHLVKTRTHALIFCKQDILLPÖNNAVLVAGGVASNFYIRRAELITNATQCTLLCP 177  
DB 275 VARHISKKHIFFEESLSEQEKLPKQ---LVIGGVANQYIFGAISKLSAAHNVTTIKVL 331  
QY 178 PRLCTDNGIMAMNGIERLRAGLGIHDIEGIRYEP-----KCPGVDISKEVG 226  
DB 332 LSLCTDNGAMTAYAGVWRVFKA--GATADL-GVSVRPRMPL 336  
QY 227 EASIKVPQPKM 237  
DB 385 ---IDTPRRKL 392

RESULT 11  
US-10-067-443-28  
; Sequence 28, Application US/10067443  
; Patent No. 6642041  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I  
; TITLE OF INVENTION: SPINAL CORD, MP-1  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 28  
; LENGTH: 421  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; US-10-067-443-28

Query Match 16.2%; Score 201.5; DB 4; Length 421;  
Best Local Similarity 27.5%; Pred. No. 6.3e-16;  
Matches 69; Conservative 48; Mismatches 99; Indels 35; Gaps 7;

QY 1 LALVGVSDPFLILGSLDIPAGMDLVKVARSLIKRPECSTMSGKAIHLAKQ-- 58  
DB 163 LLSVAEVEKFKLYGGVSSGPECIDKVAROLDL-GSEFPGIVGAVENTLASRAAD 221  
QY 59 -RHFHDIKPLHNAKNCDFSTGLQHTVDKIIMKEKEKEGIEKGQILSSADIAATVQHT 117  
DB 222 GHLRYPFLPNVPANANFDOIKGSYLANLERLRKNSTSID-----IPDFCSLDQNT 274  
QY 118 MACHLVKTRTRAILFCORDDLPPONNAVTVASGGVASFYIRALEILTNAOCTLLCP 177  
DB 275 VAHHSISKHIFPESLSEGEKLPKQ---LVIGGGAANOYIFGATSKLSAANVTITIKL 331  
QY 178 PRLCTDNGMIAMNGIERLRAGLGLHDIGIRYEP-----KCPAGVDISKEVG 226  
DB 332 LSLCTDNAMEIAVSGL-----LMLVNRSEAIWWRPNDIPPTIYAHARSDIGTASSEI 384  
QY 227 EASIKVPQLKM 237  
DB 385 ---IDTPRRKL 392

RESULT 12  
US-08-087-797-2  
Sequence 2, Application US/08087797  
Patent No. 5543312  
GENERAL INFORMATION:  
APPLICANT: Mellors, Alan  
APPLICANT: lo, Reggie Y.C.  
APPLICANT: Abdullah, Khalid M.  
TITLE OF INVENTION: Pasteurella Haemolytica  
TITLE OF INVENTION: Glycoprotease  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bell, Seitzer, Park & Gibson, P.A.  
STREET: 1211 East Morehead Street,  
CITY: Charlotte  
STATE: No. 5543312ch Carolina  
COUNTRY: United States  
ZIP: 28234  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087,797  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Layton, Jr., Samuel G.  
REGISTRATION NUMBER: 22807  
REFERENCE/DOCKET NUMBER: 3374-80  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 704 377 1561  
TELEFAX: 704 334 2014  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 325 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-087-797-2

Query Match 16.1%; Score 200; DB 1; Length 325;  
Best Local Similarity 31.6%; Pred. No. 6.5e-16;  
Matches 62; Conservative 25; Mismatches 91; Indels 18; Gaps 5;

QY 2 LALVGVSDPFLILGSLDIPAGMDLVKVARSLIKRPECSTMSGKAIHLAKQ-- 61  
DB 142 LVVVDGQYELLGSESIDDAAGEAFDKYKILGL-----AGVAMSKLASGTPNR 194  
QY 62 FDIKPLHNAKNCDFSTGLQHTVDKIIMKEKEKEGIEKGQILSSADIAATVQHTMA 121

DB 195 FRPFRMTDRPGLDFSPGLKTPAANTIRANINENELDEQ---TKCDIAHAFQAAV--- 248  
QY 122 LVKRTTRAILFCORDDLPPONNAVTVASGGVASFYIRALEILTNAOCTLLCPRLC 181  
DB 249 ---VDTIKCK-RALBGTGYKALVMAGVSANRQRLADLAEKMKLKGVEVYPRPQC 303  
QY 182 TDNGMIAMNGIERLR 197  
DB 304 TDNGMIATVGLRLK 319

RESULT 13  
US-09-107-532A-6609  
Sequence 6609, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: CTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 6609:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 363 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (8) LOCATION 1...363  
SEQUENCE DESCRIPTION: SEQ ID NO: 6609:  
US-09-107-532A-6609

Query Match 15.9%; Score 197; DB 4; Length 363;  
Best Local Similarity 32.6%; Pred. No. 1.8e-15;  
Matches 63; Conservative 34; Mismatches 64; Indels 32; Gaps 9;

QY 2 LALVGVSDPFLILGSLDIPAGMDLVKVARSLIKRPECSTMSGKAIHLAKQ-- 60  
DB 169 LVVWDEDSYRIIGETRDADAGEAYDKVGRVIGL-----SYPSGKEIDQLAHQCKDNY 221  
QY 61 HFDIKPLHNAKNCDFSTGLQHTVDKIIMK-KRKEGIEKGQILSSADIAATVQHTMA 119

Db 222 HF--PRAMIHEDNDYDFSGSKSAFINLVHNAQORGEDLDKN-----DLAASFOASVI 272  
Qy 120 CHLVKTRTRAILFKQKRDLPQNNNAV--LVASGVASNFYIRALEILITNA--TOCTLLC 175  
Db 273 DVLTITKLRA---C-----QNTVVKQLVVAAGVAAAGLREGIQAALSAALPEVELVI 322  
Qy 176 PPPRLCTDNGIMI 188  
Db 323 PPIRLCGDNAMI 335

RESULT 14  
US-10-067-443-5  
; Sequence 5, Application US/10067443  
; Patent No. 6642041  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 327  
; TYPE: PRT  
; ORGANISM: Thermotoga maritima  
US-10-067-443-5

Query Match 15.6%; Score 193.5; DB 4; Length 327;  
Best Local Similarity 28.9%; Pred. No. 4.2e-15;  
Matches 55; Conservative 38; Mismatches 68; Indels 29; Gaps 6;

Qy 13 LIGKSLDIPGMDLKVARRLSLIGHPECSWGGKALEHLAKQGNRFHDIKPLHNAK 72  
Db 151 VLGETLDSAGEAFKVARLGL-GYP-----GGPVIDRVAKKDDPEKYSFPPRLDD 203  
Qy 73 NCDPSFTGLQHTVDKIIMKEKEBEGIEKQILSSAADIATVQHTMACHLVKTRTRAILF 132  
Db 204 SYNSFPAKLG--TSVLFLQREKGYK-----VEDVASQKAVVDLVEKTR----- 249  
Qy 133 CKQKDLFPQNNAV--LVASGVASNFYIRALEILITNAQCTLLCPPRLCTDNGIMIAM 190  
Db 250 -----LARNIGIRKIAFVGVAANSMLREVRKRAERNVYVFPPLELCTDNALMVAK 303  
Qy 191 NGIERLRAGL 200  
Db 304 AGYEKAKRGM 313

RESULT 15  
US-08-987-121A-4  
; Sequence 4, Application US/08987121A  
; GENERAL INFORMATION:  
; APPLICANT: Hoskins, Jo Ann  
; APPLICANT: Tang, Joseph Chiu-Chung  
; APPLICANT: Treadway, Patti Jean  
; TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: U.S.  
; ZIP: 46285  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/987,121A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Webster, Thomas D.  
; REGISTRATION NUMBER: 39,872  
; REFERENCE/DOCKET NUMBER: X-  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3334  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 336 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULAR TYPE: protein  
US-08-987-121A-4

Query Match 15.3%; Score 190; DB 3; Length 336;  
Best Local Similarity 31.1%; Pred. No. 1.2e-14;  
Matches 61; Conservative 32; Mismatches 73; Indels 30; Gaps 8;

Qy 2 LALVGVSDPFLIGKSLDIPGMDLKVARRLSLIGHPECSWGGKALEHLAKQGNRFH 61  
Db 144 LVYVSEAGDKYIVGETRDAVGEAYDKVRWGL-----TYPAREIDELAHQHDI- 195  
Qy 62 FDIKPEPLHNAKNDPSFTGLQHTVDKIIMKEKEBEGIEKQILSSAADIATVQHTMACH 121  
Db 196 YDFPRAMIKEDNLEFSGLSKSAFINLHNAE-----QKGESIST-EDICASFQAAVMDI 249  
Qy 122 LVKTRTRAILFKQKRDLPQNNNAV--LVASGVASNFYIRALEILITNAQCTLLCPPRLC 181  
Db 250 LMAKTKKAL-----EKYPVK--TLVAVAGVANKGLRRL--ATEITVNVVILPPLRLC 299  
Qy 182 TDNGIMIA-----NN 191  
Db 300 GDNAGMIAVAVSEWN 315

Search completed: November 10, 2005, 16:54:35  
Job time : 20.2239 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2005, 16:43:58 ; Search time 63.1272 Seconds

(without alignment)  
1584.102 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414

Perfect score: 1 LHALVQGVSDPFLIGSLIDT.....DISKEVGASIKVQLKMEI 239

Sequence: 1 LHALVQGVSDPFLIGSLIDT.....DISKEVGASIKVQLKMEI 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867879 seqs, 418409474 residues

Total number of hits satisfying chosen parameters: 1867879

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09C\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US10E\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
21: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*  
22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1240	100.0	267	14	US-10-067-443-22
2	1240	100.0	267	15	US-10-649-273-22
3	1240	100.0	267	15	US-10-651-722-22
4	1240	100.0	414	14	US-10-067-443-2
5	1240	100.0	414	15	US-10-649-273-2
6	1240	100.0	414	15	US-10-651-722-2
7	1240	100.0	414	17	US-10-480-988-8
8	1240	100.0	414	17	US-10-480-988-8
9	1240	100.0	414	17	US-10-480-988-8
10	1240	100.0	414	17	US-10-480-988-8
11	1240	100.0	414	17	US-10-480-988-8
12	1240	100.0	414	17	US-10-480-988-8
13	1240	100.0	414	17	US-10-480-988-8
14	1240	100.0	414	17	US-10-480-988-8
15	1240	100.0	414	17	US-10-480-988-8
16	1240	100.0	414	17	US-10-480-988-8
17	1240	100.0	414	17	US-10-480-988-8
18	1240	100.0	414	17	US-10-480-988-8
19	1240	100.0	414	17	US-10-480-988-8
20	1240	100.0	414	17	US-10-480-988-8
21	1240	100.0	414	17	US-10-480-988-8
22	1240	100.0	414	17	US-10-480-988-8

12	960	77.4	364	15	US-10-094-749-2039	Sequence 2039, Ap
13	337	27.2	445	15	US-10-424-599-209259	Sequence 209259,
14	332.5	26.8	409	20	US-11-097-143-34191	Sequence 34191, A
15	326	26.3	463	14	US-10-067-443-3	Sequence 3, Appl
16	326	26.3	463	15	US-10-649-273-3	Sequence 3, Appl
17	326	26.3	463	15	US-10-651-722-3	Sequence 3, Appl
18	270	21.8	444	16	US-10-437-963-113732	Sequence 113732,
19	257	20.7	382	15	US-10-282-122A-50858	Sequence 50858, A
20	254	20.5	333	14	US-10-012-140-25	Sequence 25, Appl
21	248.5	20.0	348	15	US-10-282-122A-63156	Sequence 63156, A
22	247	19.9	312	15	US-10-282-122A-44499	Sequence 44499, A
23	247	19.9	343	15	US-10-282-122A-67227	Sequence 67227, A
24	239	19.3	231	14	US-10-081-051-9	Sequence 11798, A
25	239	18.6	341	9	US-09-815-242-11798	Sequence 66200, A
26	231	18.6	341	15	US-10-282-122A-66200	Sequence 66200, A
27	231	18.6	341	18	US-10-958-216-50	Sequence 52, Appl
28	229	18.5	341	18	US-10-958-216-52	Sequence 24, Appl
29	228.5	18.4	347	14	US-10-012-140-24	Sequence 52, Appl
30	224	18.1	342	9	US-09-815-242-11043	Sequence 11043, A
31	224	18.1	342	15	US-10-282-122A-58204	Sequence 58204, A
32	213.5	17.2	347	20	US-11-097-143-10128	Sequence 10128, A
33	213	17.2	337	9	US-09-815-242-13780	Sequence 13780, A
34	213	17.2	337	15	US-10-282-122A-75485	Sequence 75485, A
35	212	17.1	340	15	US-10-282-122A-68438	Sequence 68438, A
36	211	17.0	41	14	US-10-012-140-26	Sequence 26, Appl
37	210	16.9	337	15	US-10-282-122A-78161	Sequence 78161, A
38	208	16.8	335	15	US-10-282-122A-55404	Sequence 55404, A
39	207.5	16.7	341	15	US-10-282-122A-67993	Sequence 67993, A
40	207	16.7	441	16	US-10-437-963-111620	Sequence 111620,
41	206	16.6	337	9	US-09-815-242-110304	Sequence 10304, A
42	206	16.6	337	15	US-10-282-122A-56695	Sequence 56695, A
43	206	16.6	396	15	US-10-425-114-67198	Sequence 67198, A
44	201.5	16.2	421	14	US-10-067-443-4	Sequence 4, Appl
45	201.5	16.2	421	14	US-10-067-443-28	Sequence 28, Appl

## ALIGNMENTS

RESULT 1

US-10-067-443-22

Sequence 22, Application US/10067443

Publication No. US20030082782A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN

FILE REFERENCE: D0073 NP

CURRENT APPLICATION NUMBER: US/10/067,443

CURRENT FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.0

SEQ ID NO 22

LENGTH: 267

TYPE: PRT

ORGANISM: homo sapiens

US-10-067-443-22

Query Match 100.0%; Score 1240; DB 14; Length 267;

Best local Similarity 100.0%; Pred. No. 2.3e-127;

Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LHALVQGVSDPFLIGSLIDTAPCDMDKXARRSLIKHPCSTMSGSKAIENHAKGNRP 60

DB 29 LHALVQGVSDPFLIGSLIDTAPCDMDKXARRSLIKHPCSTMSGSKAIENHAKGNRP 60

QY 61 HFDIKPPLHAKKDFSPFGLQHVTDKIIIMKCKEKGIEKGQITSSAADIATVQHTMAC 120

DB 89 HFDIKPPLHAKKDFSPFGLQHVTDKIIIMKCKEKGIEKGQITSSAADIATVQHTMAC 148



```
; Sequence 2, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-649-273-2

Query Match          100.0%; Score 1240; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 4.3e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 LLAIVQGVSDFLILKSLDIAPGDMLDKVARSLTIKHECSTMSGKAI EHLAKQGNRF 60
|||||
Db 176 LLAIVQGVSDFLILKSLDIAPGDMLDKVARSLTIKHECSTMSGKAI EHLAKQGNRF 235
|||||

Cy 61 HPDIKPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 120
Db 236 HPDIKPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 295
|||||

Cy 121 HLVKRTHRAILFCCKQDILLPQNNAVLVASGVASNFYIRRALEILTNAQTCTLLCPPPRL 180
Db 296 HLVKRTHRAILFCCKQDILLPQNNAVLVASGVASNFYIRRALEILTNAQTCTLLCPPPRL 355
|||||

Cy 181 CTDNGIMIAMNGIERLRAGILGILHDI EGIRYRKPCLGVDISKEVGEASIKVPOLKMEI 239
Db 356 CTDNGIMIAMNGIERLRAGILGILHDI EGIRYRKPCLGVDISKEVGEASIKVPOLKMEI 414
|||||

RESULT 6
US-10-651-722-2
; Sequence 2, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-651-722-2

Query Match          100.0%; Score 1240; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 4.3e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 LLAIVQGVSDFLILKSLDIAPGDMLDKVARSLTIKHECSTMSGKAI EHLAKQGNRF 60
|||||
```

```
Db 176 LLAIVQGVSDFLILKSLDIAPGDMLDKVARSLTIKHECSTMSGKAI EHLAKQGNRF 235
|||||

Cy 61 HPDIKPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 120
Db 236 HPDIKPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 295
|||||

Cy 121 HLVKRTHRAILFCCKQDILLPQNNAVLVASGVASNFYIRRALEILTNAQTCTLLCPPPRL 180
Db 296 HLVKRTHRAILFCCKQDILLPQNNAVLVASGVASNFYIRRALEILTNAQTCTLLCPPPRL 355
|||||

Cy 181 CTDNGIMIAMNGIERLRAGILGILHDI EGIRYRKPCLGVDISKEVGEASIKVPOLKMEI 239
Db 356 CTDNGIMIAMNGIERLRAGILGILHDI EGIRYRKPCLGVDISKEVGEASIKVPOLKMEI 414
|||||

RESULT 7
US-10-480-988-8
; Sequence 8, Application US/10480988
; Publication No. US20050069877A1
; GENERAL INFORMATION:
; APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;
; APPLICANT: SWARNAKAR, Anita; HARALIA, April J.A.;
; APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;
; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;
; APPLICANT: HONCHELL, Cynthia D.; NGUYEN, Daniel B.;
; APPLICANT: LU, Dyung Aina M.; LEE, Ernestine A.;
; APPLICANT: YUE, Henry; FORSYTHE, Ian J.;
; APPLICANT: BARROSO, Ines; RAMKUMAR, Jayalaxmi;
; APPLICANT: GRIFPIN, Jennifer A.; LI, Doana X.;
; APPLICANT: YANG, Junming; THANGAVELOU, Kavitha;
; APPLICANT: GIETZEN, Kimberley J.; DING, Li;
; APPLICANT: BAUGHN, Maribel R.; BOROWSKY, Mark L.;
; APPLICANT: YAO, Monique G.; CHAWLA, Narinder K.;
; APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
; APPLICANT: LEE, Sally; BECHA, Shanya D.;
; APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;
; APPLICANT: ELLIOTT, Vicki S.; LUO, Wen;
; APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;
; APPLICANT: LU, Yan; ZEBARADIAN, Yeganeh
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PF-1040 USN
; CURRENT APPLICATION NUMBER: US/10/480,988
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: PCT/US02/19360
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/300,508
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/303,445
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/305,405
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/311,442
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 60/314,821
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/315,992
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 60/378,205
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7632424CD1
US-10-480-988-8

Query Match          100.0%; Score 1240; DB 17; Length 414;
Best Local Similarity 100.0%; Pred. No. 4.3e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



```
QY      1 LIALVQGVSDPFLILGKSLDIAFGDMLDKVARRLSLIKHPBCSTMSGGAIEHLAKQGNRF 60
      176 LIALVQGVSDPFLILGKSLDIAFGDMLDKVARRLSLIKHPBCSTMSGGAIEHLAKQGNRF 235
DB
QY      61 HFDIKPPLHAHAKNCDPSFTGLQHTVDKTIIMKEKEGIEKQIILSSAADIAATVQHTMAC 120
      236 HFDIKPPLHAHAKNCDPSFTGLQHTVDKTIIMKEKEGIEKQIILSSAADIAATVQHTMAC 295
DB
QY      121 HLVRTRRAILFCQORDLLPQNNAVLVASGVASNFYIRRALEILTNAQTCTLLCPPRL 180
      296 HLVRTRRAILFCQORDLLPQNNAVLVASGVASNFYIRRALEILTNAQTCTLLCPPRL 355
DB
QY      181 CTDNGIMIANNGIERLRAGLGIILHDIEGIRYEPKCPILGVDISKEVGEASIKVPOLKMEI 239
      356 CTDNGIMIANNGIERLRAGLGIILHDIEGIRYEPKCPILGVDISKEVGEASIKVPOLKMEI 414
DB
```

## RESULT 8

```
US-10-067-443-19
; Sequence 19, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-443-19
```

```
Query Match      97.9%; Score 1213.5; DB 14; Length 439;
Best Local Similarity 90.2%; Pred. No. 3.9e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;
```

```
QY      1 LIALVQGVSDPFLILGKSLDIAFGDMLDKVARRLSLIKHPBCSTMSGGAIEHLAKQGNRF 60
      176 LIALVQGVSDPFLILGKSLDIAFGDMLDKVARRLSLIKHPBCSTMSGGAIEHLAKQGNRF 235
DB
QY      61 HFDIKPPLHAHAKNCDPSFTGLQHTVDKTIIMKEKEGIEKQIILSSAADIAATVQHTMAC 120
      236 HFDIKPPLHAHAKNCDPSFTGLQHTVDKTIIMKEKEGIEKQIILSSAADIAATVQHTMAC 295
DB
QY      99 ---EKGQIILSSAADIAATVQHTMACHLVTRTRAILFCQORDLLPQNNAVLVASGVASNF 155
      296 CRYEKQIILSSAADIAATVQHTMACHLVTRTRAILFCQORDLLPQNNAVLVASGVASNF 355
DB
QY      156 FYIRRALEILTNAQTCTLLCPPRLCTDNGIMIANNGIERLRAGLGIILHDIEGIRYEPKC 215
      356 FYIRRALEILTNAQTCTLLCPPRLCTDNGIMIANNGIERLRAGLGIILHDIEGIRYEPKC 415
DB
QY      216 PLGVDISKEVGEASIKVPOLKMEI 239
      416 PLGVDISKEVGEASIKVPOLKMEI 439
DB
```

## RESULT 9

```
US-10-649-273-19
; Sequence 19, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
```

```
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-649-273-19
```

```
Query Match      97.9%; Score 1213.5; DB 15; Length 439;
Best Local Similarity 90.2%; Pred. No. 3.9e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;
```

```
QY      1 LIALVQGVSDPFLILGKSLDIAFGDMLDKVARRLSLIKHPBCSTMSGGAIEHLAKQGNRF 60
      176 LIALVQGVSDPFLILGKSLDIAFGDMLDKVARRLSLIKHPBCSTMSGGAIEHLAKQGNRF 235
DB
QY      61 HFDIKPPLHAHAKNCDPSFTGLQHTVDKTIIMKEKEGIEKQIILSSAADIAATVQHTMAC 120
      236 HFDIKPPLHAHAKNCDPSFTGLQHTVDKTIIMKEKEGIEKQIILSSAADIAATVQHTMAC 295
DB
QY      99 ---EKGQIILSSAADIAATVQHTMACHLVTRTRAILFCQORDLLPQNNAVLVASGVASNF 155
      296 CRYEKQIILSSAADIAATVQHTMACHLVTRTRAILFCQORDLLPQNNAVLVASGVASNF 355
DB
QY      156 FYIRRALEILTNAQTCTLLCPPRLCTDNGIMIANNGIERLRAGLGIILHDIEGIRYEPKC 215
      356 FYIRRALEILTNAQTCTLLCPPRLCTDNGIMIANNGIERLRAGLGIILHDIEGIRYEPKC 415
DB
QY      216 PLGVDISKEVGEASIKVPOLKMEI 239
      416 PLGVDISKEVGEASIKVPOLKMEI 439
DB
```

## RESULT 10

```
US-10-651-722-19
; Sequence 19, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-651-722-19
```

```
Query Match      97.9%; Score 1213.5; DB 15; Length 439;
Best Local Similarity 90.2%; Pred. No. 3.9e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;
```

```
QY      1 LIALVQGVSDPFLILGKSLDIAFGDMLDKVARRLSLIKHPBCSTMSGGAIEHLAKQGNRF 60
      176 LIALVQGVSDPFLILGKSLDIAFGDMLDKVARRLSLIKHPBCSTMSGGAIEHLAKQGNRF 235
DB
```

```
QY 61 HFDIKPPLHAKNCPSFTGLQHTVDKIIMKKEKEGI----- 98
D 236 HFDIKPPLHAKNCPSFTGLQHTVDKIIMKKEKEGIPLISKEQINIPGLCLKIAHF 295
QY 99 ---EKGOILSSADIAATQHTMACHLVKTRTHAILFCRKORDLLPNNNAVTVASGVASN 155
D 296 CRYEKQILSSADIAATVQHTMACHLVKTRTHAILFCRKORDLLPNNNAVTVASGVASN 355
QY 156 FYIRRALIELTNATQCTLLCPPRLCTDNGIMIANWGIERLRLAGLILHDIEGIRYEPKC 215
D 356 FYIRRALIELTNATQCTLLCPPRLCTDNGIMIANWGIERLRLAGLILHDIEGIRYEPKC 415
QY 216 PLGVDISKEVGEASIKVPOLKMEI 239
D 416 PLGVDISKEVGEASIKVPOLKMEI 439

RESULT 11
US-10-012-140-5
; Sequence 5, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leibyl, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USSES
; TITLE OF INVENTION: THEROOP
; FILE REFERENCE: 381552004500
; CURRENT APPLICATION NUMBER: US/10/012,140
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-140-5

Query Match 97.0%; Score 1203; DB 14; Length 414;
Best Local Similarity 97.1%; Pred. No. 5,1e-123;
Matches 232; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 LLAIVGVSDPFLILGKSLDIPGMDLVKARRLSLIKHPECSTMSGKAI EHLAKGNRF 60
D 176 LLAIVGVSDPFLILGKSLDIPGMDLVKARRLSLIKHPECSTMSGKAI EHLAKGNRF 235
QY 61 HFDIKPPLHAKNCPSFTGLQHTVDKIIMKKEKEGIKGOILSSADIAATVQHTMAC 120
D 236 HFDIKPPLHAKNCPSFTGLQHTVDKNNENKQEBGIEKGOILSSADIAATVQHTMAC 295
QY 121 HLVKTRTHAILFCRKORDLLPNNNAVTVASGVASNYIRRALIELTNATQCTLLCPPRL 180
D 296 HLVKTRTHAILFCRKORDLLPNNNAVTVASGVASNYIRRALIELTNATQCTLLCPPRL 355
QY 181 CTDNGIMIANWGIERLRLAGLILHDIEGIRYEPKCPLGVDISKEVGEASIKVPOLKMEI 239
D 356 CTDNGIMIANWGIERLRLAGLILHDIEGIRYEPKCPLGVDISKEVGEASIKVPOLKMEI 414

RESULT 12
US-10-094-749-2039
; Sequence 2035, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
```

```
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2039
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2039

Query Match 77.4%; Score 960; DB 15; Length 364;
Best Local Similarity 98.4%; Pred. No. 2,2e-96;
Matches 186; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 LLAIVGVSDPFLILGKSLDIPGMDLVKARRLSLIKHPECSTMSGKAI EHLAKGNRF 60
D 176 LLAIVGVSDPFLILGKSLDIPGMDLVKARRLSLIKHPECSTMSGKAI EHLAKGNRF 235
QY 61 HFDIKPPLHAKNCPSFTGLQHTVDKIIMKKEKEGIKGOILSSADIAATVQHTMAC 120
D 236 HFDIKPPLHAKNCPSFTGLQHTVDKIIMKKEKEGIKGOILSSADIAATVQHTMAC 295
QY 121 HLVKTRTHAILFCRKORDLLPNNNAVTVASGVASNYIRRALIELTNATQCTLLCPPRL 180
D 296 HLVKTRTHAILFCRKORDLLPNNNAVTVASGVASNYIRRALIELTNATQCTLLCPPRL 355
QY 181 CTDNGIMIA 189
D 356 CTDNGIMIA 364

RESULT 13
US-10-424-598-209259
; Sequence 209259, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 209259
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Glycine max
```

FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_3098C.1.pcp  
US-10-424-599-209259

Query Match 27.2%; Score 337; DB 15; Length 445;  
Best Local Similarity 38.5%; Pred. No. 8,9e-28;  
Matches 85; Conservative 29; Mismatches 83; Indels 24; Gaps 5;

QY 1 LIALVGVSDPFLIGKSLDIAPGMDLVKVARSLIKHPECSTMSGKAIHIAKQGNRF 60  
DB 194 LVLVARDIGQYIQLGTTTIDDAIGEAIDTKAKMGLDLR-----RSGGPAIEKLAIEGNAE 248  
QY 61 HFDIKPPLHAKNCDSEFTGLQ----HYTDKIIMKEKEGIEKQILSSAADIATVQ 115  
DB 249 SVKFSIPMKQKDCNFSYAGLKTQVRLAIESKKIDAKIPISSANGDRU-SRADIAASQ 307  
QY 116 HTMACHLVKTRHAILFCQKQDLPQNNNAVVASGVASNFYIRALBITLTAQCTLLC 175  
DB 308 RIAVLHLEERGERAIOMLAKMEPSIRH--LVVSGGVASNQYRARLDVVKKQGLQVLC 364  
QY 176 PPRLCTDNGIMIAMNGIERLRAGIGILHDIIEGIRYEPKCP 216  
DB 365 PPRLCTDNGVIAWTGIEHFRMG-----RYDPPPP 395

RESULT 14  
US-11-097-143-34191  
Sequence 34191, Application US/11097143  
Publication No. US20050208558A1  
GENERAL INFORMATION:  
APPLICANT: Venter, J. Craig  
TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
FILE REFERENCE: CLO00728  
CURRENT APPLICATION NUMBER: US/11/097,143  
CURRENT FILING DATE: 2005-04-04  
PRIOR APPLICATION NUMBER: 60/157,832  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: 60/160,191  
PRIOR FILING DATE: 1999-10-19  
PRIOR APPLICATION NUMBER: 60/161,932  
PRIOR FILING DATE: 1999-10-28  
PRIOR APPLICATION NUMBER: 60/164,769  
PRIOR FILING DATE: 1999-11-12  
PRIOR APPLICATION NUMBER: 60/173,383  
PRIOR FILING DATE: 1999-12-28  
PRIOR APPLICATION NUMBER: 60/175,693  
PRIOR FILING DATE: 2000-01-12  
PRIOR APPLICATION NUMBER: 60/184,831  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/191,637  
PRIOR FILING DATE: 2000-03-23  
NUMBER OF SEQ ID NOS: 43008  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 34191  
LENGTH: 409  
TYPE: PRT  
ORGANISM: DROSOPHILA  
US-11-097-143-34191

Query Match 26.8%; Score 332.5; DB 20; Length 409;  
Best Local Similarity 35.9%; Pred. No. 2.5e-27;  
Matches 83; Conservative 38; Mismatches 97; Indels 13; Gaps 5;

QY 2 LALVGVSDPFLIGKSLDIAPGMDLVKVARSLIKHPECSTMSGKAIHIAKQGNRF 60  
DB 167 LVLVANGPRLTLGLGTLDDAPGAPDKIGRLRLHLTPYRLMNGRAIEHAAQLASDP 226  
QY 61 HFDIKPPLHAKNCDSEFTGLQHTVDKIIMKEKEGIEKQILSSAADIATVQHTMAC 120  
DB 227 AVFPLPLAQGNKCNFSPAGIKGNRPAIRABERARTPPDGVISNYGDFCGLASVSR 286

QY 121 HLVKTRHAILFC--KQDRLPQNNNAVVASGVASNFYIRALBITLTAQCTLLCPPP 178  
DB 287 HLMHRTQRAIEYCLPBRQLFGDTPPTLVMSGGVANNDAIVANIEHLAAQYGRSFRPSK 346  
QY 179 RCTDNGIMIAMNGIERLRAGIGILHDI-GIRYEPKCLGVDISVEGEA 228  
DB 347 RYCSBNGVIAWHGVEOL-----LQDKASTRYDYD---SIDIQSAGFA 388

RESULT 15  
US-10-067-443-3  
Sequence 3, Application US/10067443  
Publication No. US20030082782A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED I  
FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
CURRENT FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 3  
LENGTH: 463  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-10-067-443-3

Query Match 26.3%; Score 326; DB 14; Length 463;  
Best Local Similarity 33.6%; Pred. No. 1.5e-26;  
Matches 82; Conservative 34; Mismatches 84; Indels 44; Gaps 5;

QY 1 LIALVGVSDPFLIGKSLDIAPGMDLVKVARSLIKHPECSTMSGKAIHIAKQGNRF 60  
DB 221 LVLVANKLQYIQLGTTTIDDAIGEAIDTKAKMGLDMH-----RSGGPAVEELALBEDAK 275  
QY 61 HFDIKPPLHAKNCDSEFTGLQHTVDKIIMKEKEGIEKQILSSAADIATVQHTMAC 120  
DB 276 SVKFNVPKMKYKDCNFSYAGLKTQVRLAIEAK-----IRNRADIAASFORVAVL 325  
QY 121 HLVKTRHAILFCQKQDLPQNNNAVVASGVASNFYIRALBITLTAQCTLLCPPRL 180  
DB 326 HLEEKCEAIDWALE--LEPSIKHNVISGVASNRYRLNNIVENKNKLKLVCPPSL 382  
QY 181 CTDNGIMIAMNGIERLRAGIGILHDIIEGIRY-----PKCPGVDISKE 224  
DB 383 CTDNGVIAWTGIEHFRVG-----RYDPPPAIEPEDYVYDLRPFVIGEEYAKG 432  
QY 225 VGEA 228  
DB 433 RSEA 436

Search completed: November 10, 2005, 16:58:52  
Job time : 64.1272 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: November 10, 2005, 16:52:13 ; Search time 4714.79 Seconds  
(without alignments)  
2456.267 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414

Perfect score: 1240  
Sequence: 1 LLLLVGVSDFLILGKSLDI.....DISKVGASIKVPLAKMEI 239

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 2427607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n model -DEV=xlp  
-Q/cgr2/USFTO.spool/p/US10649273/runat\_02112005\_091338\_15550/app.query.fasta\_1.1429  
-DB=Genemb1 -GPM=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blowsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273 @CNC 1.1 7742 @runat\_02112005\_091338\_15550 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGESUBSTRY -NEG\_SCORES=0 -WAIT -DSPBLCK=100 -LONGLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmb1.\*

1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1240	100.0	1416	AR541929	Sequence
2	1240	100.0	1526	AR428809	Sequence
3	1240	100.0	1908	BC011904	Homo sapi
4	1240	100.0	2197	AR428803	Sequence

5	1213.5	97.9	1387	6	AR428808	Sequence
6	1213.5	97.9	1387	9	HS4295148	Sequence
7	1203	97.0	1245	6	AX646697	Sequence
8	1203	97.0	1820	6	AX646695	Sequence
9	1086	87.6	1844	10	BC058172	Mus muscu
10	1079	87.0	1017	10	BC038910	Mus muscu
11	1059	85.4	2208	6	AK713716	Sequence
12	1059	85.4	2208	9	AK055441	Homo sapi
13	983	79.3	1546	10	BC078974	Rattus no
14	944.5	76.2	1109	10	BC051211	Mus muscu
15	809	65.2	860	5	BX930694	Gallus ga
16	806	65.0	1522	5	BX934991	Gallus ga
17	803	64.8	1558	6	BX930963	Gallus ga
18	725	58.5	14364	6	AR428807	Sequence
19	725	58.5	84115	9	AC013468	Homo sapi
20	666	53.7	249601	2	AC114153	Rattus no
21	666	53.7	308652	2	AC121478	Rattus no
22	646.5	52.1	256751	10	AC122925	Mus muscu
23	610	49.2	875	6	CQ721898	Sequence
24	384.5	31.0	1474	3	AK113378	Ciona int
25	369.5	29.8	121251	5	AL591593	Zebrafish
26	356	28.7	20021	10	AF367967	Mus muscu
27	356	28.7	179252	10	AF131205	Mus muscu
28	345	27.8	117322	5	AL672217	Zebrafish
29	332.5	26.8	1576	3	AY051882	Drosophila
30	332.5	26.8	1601	6	CQ606432	Sequence
31	332.5	26.8	3656	6	CQ606431	Sequence
32	332.5	26.8	14679	2	AC018262	Drosophila
33	332.5	26.8	180263	3	AC010671	Drosophila
34	332.5	26.8	207432	3	AE003513	Drosophila
35	324	26.1	1443	8	AY024338	Arabidops
36	324	26.1	1474	8	AY117283	Arabidops
37	324	26.1	1567	8	AY063864	Arabidops
38	322	26.0	1557	8	AY084577	Arabidops
39	311.5	25.1	1672	8	AK070912	Oryza sat
40	305	24.6	110000	2	AP006501	Continuation (9 of
41	262.5	21.2	333800	1	SM591792	Stenohizo
42	260	20.0	1718	8	AK099965	Oryza sat
43	257	20.7	349354	1	BX640416	Bordetella
44	254	20.5	260271	1	AE017258	Molbachia
45	253	20.4	349116	1	AP003003	Mesorhizo

#### ALIGNMENTS

RESULT 1	AR541929	AR541929	1416 bp	DNA	linear	PAT 08-OCT-2004
LOCUS	AR541929	Sequence 177 from patent US 6743619.				
DEFINITION	AR541929					
ACCESSION	AR541929.1	GI:53934009				
VERSION	AR541929.1					
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 1416)					
AUTHORS	Tang, Y.T., Zhou, P., Goodrich, R., Liu, C., Asundi, V., Ren, F., Zhang, J., Zhao, Q.A., Yang, Y., Xue, A.J., Wehrman, T., Wang, J.-R., Wang, D., and Drmanac, R.T.					
TITLE	Nucleic acids and polypeptides					
JOURNAL	Patent: US 6743619-A 177 01-JUN-2004;					
FEATURES	Location/Qualifiers					
source	1..1416					
ORIGIN	/organism="unknown"					
	/mol_type="genomic DNA"					

Alignment Scores: 2,24e-113 Length: 1416  
Pred. No.: 1240.00 Matches: 239  
Score: 1240.00 Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 100.00%

DB: 6 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AR428809 (1-1416)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20

DB 586 CTGTGGCATTTAGTTCAAGAGGTTTCAGATTTTCTGCTTGTGAAAGCTTTTGAACATA 645

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40

DB 646 GCACCAAGTGCACAGCTTGCACAGAGTGCAGAGACCTTTCTTAATAAATCCACAG 705

QY 41 CysSerTherMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60

DB 706 TGTCTCCACCATGAGTGGTGGAAAGCCATAGAACCTTTGGCCAAACAGAAATAGATT 765

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80

DB 766 CATTTTGCATCAAACTCCCTTGATCATGCTAAAAATTTGGATTCTTTTACTGGA 825

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100

DB 826 CTTCAACAGCTTACTGATTAATAATAATGAAGAAAGAAAGAGAGATTTAGAAAG 885

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaIleValGlnHisThrMetAlaCys 120

DB 886 GGGCAATTCCTGCTTCCAGACAGACATGCTGCCACATGACACACATGAGCATGT 945

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140

DB 946 CATCTTGGAAAGAACACATCGGGCTATTCTGTTTGTAAAGCAGAGACATTTGACT 1005

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAspPheThrIleArgArg 160

DB 1006 CAAATATATGACATGCTGTCATCTGTGTGACAGATTAATCTTAATCCGCAGA 1065

QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180

DB 1066 GCTCTGAAATTTTAAACCAACGACACATGCTGCTGTCTGCTCTCCACAGATA 1125

QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200

DB 1126 TGCACTGATATGCGATTATGATTGATGCAATGATTAAGAAAGACTAGCTGCTGCT 1185

QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyArgLysCysProLeuGlyValAsp 220

DB 1186 GGCATTTTTACATGACATGAAAGGCATCCGCTATGAAACCAAAATGTCTCTTGAGATGAGC 1245

QY 221 IleSerLysGluValGlyLysValAlaSerIleLysValProGlnLeuLysMetGluIle 239

DB 1246 ATATCAAAAAGAGTTGGAGAGCTTCCATTAAGTACCAATTAATAAATGAGATTA 1302

RESULT 2

AR428809 1526 bp DNA linear PAT 18-DEC-2003

LOCUS Sequence 23 from patent US 6642041.

ACCESSION AR428809

VERSION AR428809.1 GI:40188595

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1526)

AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.

TITLE Polynucleotides encoding a novel metalloprotease, MP-1

JOURNAL Patent: US 6642041-A 23 04-NOV-2003;

FEATURES

source location/Qualifiers

1..1526

/organism="unknown"

/mol\_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.: 2,45e-113 Length: 1526

Score: 1240.00 Matches: 239

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 6 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AR428809 (1-1526)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20

DB 85 CTGTGGCATTTAGTTCAAGAGGTTTCAGATTTTCTGCTTGTGAAAGCTTTTGAACATA 144

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40

DB 145 GCACCAAGTGCACAGCTTGCACAGAGTGCAGAGACCTTTCTTAATAAATCCACAG 204

QY 41 CysSerTherMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60

DB 205 TGTCTCCACCATGAGTGGTGGAAAGCCATAGAACCTTTGGCCAAACAGAAATAGATT 264

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80

DB 265 CATTTTGCATCAAACTCCCTTGATCATGCTAAAAATTTGGATTCTTTTACTGGA 324

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100

DB 325 CTTCAACAGCTTACTGATTAATAATAATGAAGAAAGAAAGAGATTTAGAAAG 384

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaIleValGlnHisThrMetAlaCys 120

DB 385 GGGCAATTCCTGCTTCCAGACAGACATGCTGCCACAGTACACACCAATGCGATGT 444

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140

DB 445 CATCTTGGAAAGAACACATCGGGCTATTCTGTTTGTAAAGCAGAGACATTTGACT 504

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAspPheThrIleArgArg 160

DB 505 CAAATATATGACATGCTGTCATCTGTGTGACAGATTAATCTTAATCCGCAGA 564

QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180

DB 565 GCTCTGAAATTTTAAACCAACGACACATGCTGCTGTGTGCTCTCTCCACAGATA 624

QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200

DB 625 TGCACTGATATGCGATTATGATTGATGCAATGATTAAGAAAGACTAGCTGCTGCT 684

QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyArgLysCysProLeuGlyValAsp 220

DB 685 GGCATTTTTACATGACATGAAAGGCATCCGCTATGAAACCAAAATGTCTCTTGAGATGAGC 744

QY 221 IleSerLysGluValGlyLysValAlaSerIleLysValProGlnLeuLysMetGluIle 239

DB 745 ATATCAAAAAGAGTTGGAGAGCTTCCATTAAGTACCAATTAATAAATGAGATTA 801

RESULT 3

BC011904 1908 bp mRNA linear PRI 23-DEC-2003

LOCUS Homo sapiens O-sialoglycoprotein endopeptidase-like 1, mRNA (cDNA

DEFINITION clone MGC:20293 IMAGE:4121450), complete cds.

ACCESSION BC011904

VERSION BC011904.2 GI:40225818

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1908)

AUTHORS Strausberg,R.L., Feingold,B.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marinina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stappleton, M., Soares, M.B., Bonaldi, M.P., Casavant, T.L., Schaefer, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Cinci, P., Prange, C., Raha, S.S., Loguvallo, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McWhan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Wotley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Feltus, J., Helton, E., Kettman, M., Madan, A.C., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butler, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E., Schermer, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 1908)

Strausberg, R.

Direct Submission

Submitted (30-UTL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIA-MGC Project URL: <http://mgc.nci.nih.gov>

On Dec 19, 2003 this sequence version replaced gi:15080281.

Contact: MGC help desk

Email: [gcgaps-remail.nih.gov](mailto:gcgaps-remail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNI)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)

Ahler, N., Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, D., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, B., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantirip, S., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Weherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LILNI at: <http://image.llnl.gov>

Series: IRAL Plate: 28 Row: 1 Column: 22.

Location/Qualifiers

1. 1908

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="MGC:20293 IMAGE:4121450"

/tissue\_type="Muscle, rhabdomyosarcoma"

/clone\_id="NIH\_MGC\_17"

/lab\_host="DH10B-R"

/note="Vector: pOTB7"

1. 1908

/gene="OSGEP1"

/db\_xref="LOCUSID:64172"

104. 1348

/gene="OSGEP1"

/codon\_start=1

/product="OSGEP1 protein"

/protein\_id="AAH1904.1"

/db\_xref="GI:15080282"

/db\_xref="LOCUSID:64172"

/translation="MLILITKAGVFPKPKKRYEPLRSFNFHPTLPLAKIVLGIETSCDDTAAYVDENGVAGBAISQTEVHLTKGIVPPAAQOLHEENIORIVGBALSNASGVSPDSLSAIAITTIKPGIALSLGVLSPSLQVLQGLKKPFIPIHMEHALTITLNTK

misc\_feature

215. 1111

/gene="OSGEP1"

/note="Peptidase M22; Region: Glycoprotease family"

/db\_xref="CD:PFam0814"

ORIGIN

Alignment Scores:

Pred. No.: 3,19e-113 Length: 1908

Score: 1240.00 Matches: 239

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 9 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x BC011904 (1-1908)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlySerLeuAsp11e 20

DB 629 CTTGGCATTTAGTTCAAGAGTTTCAGATTTCTGCTTGGAAAGCTTTGGACATA 688

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuLysHisProGlu 40

DB 689 GCACCAAGGAGCATGCTTGCACAGGTGGCAAGAGCTTTCTTAATAAACAATCCAGAG 748

QY 41 CysSerThrMetSerGlyGlyLysAlaGlnGlnHisLeuAlaGlnGlnLysArgPhe 60

DB 749 TCCTCACCATGAGTGGTGGAGAACCATPAGAACTTTGGCCAAACAGAAATAGATT 808

QY 61 HisPheAspLysLeuProPheLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80

DB 809 CATTTGACATCAACCTCCCTTGCATCATGCTTAAATGATTTCTTTTACTGCA 868

QY 81 LeuGlnHisValThrAspLysLysIleIleMetLysGlyLysGlnGlnGlyLysGln 100

DB 869 CTTCAACAGCTTACTGATTAATAATATGAAAAAGAGAGAGCTTTGAGAG 928

QY 101 GlnGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120

DB 929 GGGCAATCTGCTTTCAGCAGACATGCTGCGACAGTACAGACCAATGCGATCT 988

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysGlnArgAspLeuPro 140

DB 989 CATCTTGGAAAAAGAACCATGCGCTATTCGTTTGTAGAGAGAGACATTTGACCT 1048

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAspPheThrLeuArg 160

DB 1049 CAATAATAGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1108

QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180

DB 1109 GCTTGGAATTTTAAACAACCAACACAGTGCATTTTGTGCTCTCTCCAGACTA 1168

QY 181 CysThrAspAspGlyIleMetIleAlaArgAspGlyIleGlnArgLeuArgAlaGlyLeu 200

DB 1169 TGCACTGATTAATGACATTAATGATGATGATGATGATGATGATGATGATGATGAT 1228

QY 201 GlnIleLeuHisAspIleGlnGlyLysLeuArgGlyArgProLysCysProLeuGlyValAsp 220

DB 1229 GGCATTTTACATGACATAGAGAGAGCATCCGCTTGAACCAAAATGCTCTTGTAGAGAGAC 1288

QY 221 IleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 239

DB 1289 ATATCAAAAGAGATTGAGAGAGCTTCCATTAAGTACCAATTAAATTAAGAGATTA 1345

RESULT 4

AR428803 2197 bp DNA linear PAT 18-DEC-2003

LOCUS

DEFINITION Sequence 1 from patent US 6642041.

ACCESSION AR428803  
VERSION AR428803.1 GI:40188589  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 2197)  
AUTHORS Chen, J., Feder, J.N., Nelson, T.C., Krystek, S.R. and Duclos, F.  
TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 6642041-A 1 04-NOV-2003;  
FEATURES  
source  
1..2197  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Alignment Scores:  
Pred. No.: 3,77e-113 Length: 2197  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AR428803 (1-2197)

QY 1 LeuleuAlaleuValGInGlyValSerApPheLeuLeuGlyLeuSerLeuAspIle 20  
DB 756 CTGTGGCATTAAGTTCAGAGAGTTTCAGATTTTCGCTTCTTGAAAGCTTTGGACATA 815

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
DB 816 GCACCAAGTGCATGCTTGCACAGGTGGCAAGACCTTTCTTATATAACATCCAGAG 875

QY 41 CysSerTherMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
DB 876 TGCCTCCACCATGAGTGGGGAAGCCATAGAACATTTGGCCAAACAGGAATGATTT 935

QY 61 HisPheAspIleLysProProLeuHisAlaLysAsnCyAspPheSerPheTherGly 80  
DB 936 CATTTTACATCAACCTCCCTTGCATCATGCTTAAATTTGATTTTCTTTTACTGGA 995

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100  
DB 996 CTTCAACAGCTTATGATTAATAATATGAAGAAAGAAAGAGATTTGAGAG 1055

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisTherMetAlaCys 120  
DB 1056 GGGCAAAATCCGCTCTTCAGCAGACATTCCTGCCACAGTACAGACACATGCGCATGT 1115

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140  
DB 1116 CATCTTGTGAAGAAACACATCGGGCTATTCGTTTGTAAAGACAGACATGTTTACTT 1175

QY 141 GlnAsnAsnAlaValIleValAlaIleSerGlyGlyValAlaIleSerAsnPheThrIleArgArg 160  
DB 1176 CAATAATATGACAGTACTGCTGTCATCTGTGTGCGCAAGTATCTTATATCCGACAG 1235

QY 161 AlaLeuGluIleLeuThrAsnAlaIleThrGlnCysThrLeuLeuCysProProProArgLeu 180  
DB 1236 GCTCTGAAATTTTAAACAAACGCAACAGTGCATTTGTGTGTCTCTCCACAGACTA 1295

QY 181 CysThrAspAsnGlyIleMetIleAlaIleThrAsnGlyIleGluArgLeuArgAlaIleLeu 200  
DB 1296 TGCATATATATGCGCATTTATATGTCATGGAATGATTTGAAGACTACGTCGCTGCTG 1355

QY 201 GllIleLeuHisAspIleGluGlyIleArgGlyArgIleProLysCysProLeuGluGlyValAsp 220  
DB 1356 GGCATTTTACATGACATAGAGGCACTCCGCTATGACCAAAATGTCCTCTTGAGATGAC 1415

QY 221 ILeSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239  
DB 1416 ATATCAAAAGAGATTTGAGAGCTTCATATAAAGTACCAATTTAAATATGAGATTA 1472

RESULT 5  
LOCUS AR428808 1387 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 21 from patent US 6642041.  
ACCESSION AR428808  
VERSION AR428808.1 GI:40188594  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1387)  
AUTHORS Chen, J., Feder, J.N., Nelson, T.C., Krystek, S.R. and Duclos, F.  
TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 6642041-A 21 04-NOV-2003;  
FEATURES  
source  
1..1387  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Alignment Scores:  
Pred. No.: 9,37e-111 Length: 1387  
Score: 1213.50 Matches: 238  
Percent Similarity: 90.15% Conservative: 0  
Best Local Similarity: 90.15% Mismatches: 1  
Query Match: 97.86% Indels: 25  
DB: 6 Gaps: 1

US-10-649-273-2\_COPY\_176\_414 (1-239) x AR428808 (1-1387)

QY 1 LeuleuAlaleuValGInGlyValSerApPheLeuLeuGlyLysSerLeuAspIle 20  
DB 549 CTGTGGCATTAAGTTCAGAGAGTTTCAGATTTTCGCTTCTTGAAAGCTTTGGACATA 608

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
DB 609 GCACCAAGTGCATGCTTGCACAGGTGGCAAGACCTTTCTTATATAAACATCCAGAG 668

QY 41 CysSerTherMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
DB 669 TGCCTCCACCATGAGTGGGGAAGCCATAGAGCATTTGGCCAAACAGGAATGATTT 728

QY 61 HisPheAspIleLysProProLeuHisAlaLysAsnCyAspPheSerPheTherGly 80  
DB 729 CATTTTACATCAACCTCCCTTGCATCATGCTTAAATTTGATTTTCTTTTACTGGA 788

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIle----- 98  
DB 789 CTTCAACAGCTTATGATTAATAATATGAAGAAAGAAAGAGAGATTAATTTCTTA 848

QY 98 ----- 98  
DB 849 ATTAGTAAAGTTGAACAGATAATATCTCGATTGTCCTTAAATAATAGCTCATTTTC 908

QY 99 -----GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGln 115  
DB 909 TGCAGTATAGAAAGGGCAAAATCTGTCTTTCAGCAGCAGACATTCCTGCAACAGTACAG 968

QY 116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135  
DB 969 CACACATGAGATGTCATCTTGTGAAGAAACACATCGGGCTATTCGTTTGTAAAGCAG 1028

QY 136 ArgAspLeuLeuProGlnAsnAsnAlaValIleValAlaIleSerGlyGlyValAlaIleSerAsn 155  
DB 1029 AGAGACTTGTACCTCAAAATATATGACATGCTGTCATCTGTGTGTGCGCAAGTAAAC 1088

QY 156 PheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaIleThrGlnCysThrLeuLeuCys 175  
DB 1089 TTCTATATCGAGAGCTCTGGAATTTTAAACAAACGCAACAGTGCATTTGTGTGT 1148

QY 176 ProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleThrAsnGlyIleGluArg 195



Db 1149 CCTCCTCCAGACTATGATGATTAATGATTTGATGATGATGATTTGATAAACA 1208  
Qy 196 leuargalagileuglylleuuhisaspillegluglylleargtyrgluprolyscys 215  
Db 1209 CTACGTGGTGGCTTGGGCAATTTTACATGACATAGAGGCATCCGCTTGAACCAAAATGT 1268  
Qy 216 ProleuglyValaspilleglylleuglylleuglylleuglylleuglylleuglylle 235  
Db 1269 CCTCTTGAGTAGATGATTAACAAAGAGTGGAGAGCTTCATTAAGTACCAAAATTA 1328  
Qy 236 lysmetglutle 239  
Db 1329 AAAATGGAGATA 1340  
RESULT 6  
LOCUS HSA295148 1387 bp mRNA linear PRI 30-OCT-2000  
DEFINITION Homo sapiens mRNA for putative sialoglycoprotease type 2.  
ACCESSION AJ295148  
VERSION AJ295148.1 GI:11071726  
KEYWORDS metalloproteinase; sialoglycoprotease.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS 1 Chen J.M., Fortunato M. and Barrett A.J.  
TITLE Cloning and sequencing of a second human putative  
sialoglycoprotease homologue  
JOURNAL 2 (bases 1 to 1387)  
AUTHORS Chen J.M.  
TITLE Direct Submission  
JOURNAL Submitted (27-OCT-2000) Chen J.M., MRC Molecular Enzymology  
Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT,  
UNITED KINGDOM  
FEATURES  
source  
1. .1387  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/tissue\_type="uterus"  
24. .1343  
/function="metalloproteinase of family M22"  
/codon\_start=1  
/product="putative sialoglycoprotease type 2"  
/protein\_id="CAC14666.1"  
/db\_xref="GI:11071727"  
/db\_xref="GOA:Q9H4B0"  
/translation="MLITKAGVFPKDSKRVLEPLSPFPHETLPIHKIVGIET  
SCDDTAAYVDENGVGENIHSGTVEHLTKGIVPPAQQGLHENTRITVOEALSSAS  
GVSPSDLSAIAITTKPGALSLGVSLPSLDLVQDKKPIPIHMEHALTIRLTK  
VEPFVLILISGHCILALVOGVSDPLILASDLIAPDMDLKVARSLIKHEPCT  
MSGGKAIEHLAKGNRPHPDIKPLHLAKNDFSTGLOHYTDKIIMKEKEGFI  
SKGQINIPGLCKLIAHPCFKYKGGIISAAIDAAVQHMACHLVKTRATLIFCK  
ORDLLPNNAVLVASGVASNFYIRBALLETITNAQCCLCPRLCTCDNIGIMANG  
IERLRGGILHIDIEGRYEPKCPGLVDISKEVBASIKVQDLKMEI"  
ORIGIN  
Alignment Scores:  
Pred. No.: 9.37e-111 length: 1387  
Score: 1213.50 Matches: 238  
Percent Similarity: 90.15% Conservatave: 0  
Best Local Similarity: 90.15% Mismatches: 1  
Query Match: 97.86% Indels: 25  
Gaps: 1  
US-10-649-273-2\_COPY\_176\_414 (1-239) x HSA295148 (1-1387)  
Qy 1 leuenu1aleuValGInglyValSerAspPheLeuLeuGlyLeuSerLeuaspille 20  
Db 549 CTGTGGCATTAAGTTCAGAGAGTTCAGATTTTCGCTTCTTGAGAAATCTTTGACATA 608

Qy 21 AlaProGlyAspMetLeuaspIysValAlaArgArgLeuSerLeuIleYshisProGlu 40  
Db 609 GCACCAAGGTGACATGCTTGCACAAAGGTGGCAAAAGACTTTCTTAATAAAATCCAGAG 668  
Qy 41 CysSerThrMetSerGlyGlyIysValAlaIleGluIshLeuAlaYsgInIysAsnArgPhe 60  
Db 669 TGCTCCACCATAGAGGTGGGAAAGCCATAGAGCATTTGGCCAAACAAAGAAATAGATT 728  
Qy 61 HisPheAspIleLeuProProLeuuhisAlaIysAsnCysAspPheSerPheThrGly 80  
Db 729 CATTTTGACATCAAACTCCCTTGATCATGCTAAAAATTTGATTTCTTTTACTGGA 788  
Qy 81 leuGlnhisValThrAspIysIleIleuetylysgIuLyugIuglylle----- 98  
Db 789 CTTCAACACGTTACTGATTAATAATATGAAAAAGAAAGAAAGAGAGATATATTCTA 848  
Qy 98 ----- 98  
Db 849 ATTAGTAAAGTTGAAACAGATTAATATCTCGATGTTGCTTAATAATAGCTGCTATTTC 908  
Qy 99 -----GluIysGlyGlnIleuSerSerAlaAlaAspIleAlaIleThrValGln 115  
Db 909 TGCAGGTATGAGAAAGGGGCAAAATCTGTCTTACAGACAGACATTTGCTCCACAGTACG 968  
Qy 116 HisThrMetAlaCysHisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGln 135  
Db 969 CACACAAATGGCATGTCATCTTGTGAAGAAACACATCGGCTATTCTGTTTGTAAAGCG 1028  
Qy 136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAsnAsn 155  
Db 1029 AGAGACTTGTATCCCAAAATATATGACATGCTGTCATCTGCTGCTGCTCCCAAGTAA 1088  
Qy 156 PheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGluCysThrLeuLeu 175  
Db 1089 TTTATATCCGAGAGCTCTGAAATTTTAAACAAAGCAACACAGTGCCTTGTGTGT 1148  
Qy 176 PropProArgLeuCysThrAspAsnGlyIleuetylleAlaIleAsnGlyIleGluArg 195  
Db 1149 CCTCCTCCAGCATATGACATGATTAATGATGATTTGATGATGATGATTTGAAAGA 1208  
Qy 196 leuargalagileuglylleuuhisaspillegluglylleargtyrgluprolyscys 215  
Db 1209 CTACGTGGTGGCTTGGGCAATTTTACATGACATAGAGGCATCCGCTTGAACCAAAATGT 1268  
Qy 216 ProleuglyValaspilleglylleuglylleuglylleuglylleuglylleuglylle 235  
Db 1269 CCTCTTGAGTAGATGATTAACAAAGAGTGGAGAGCTTCATTAAGTACCAAAATTA 1328  
Qy 236 lysmetglutle 239  
Db 1329 AAAATGGAGATA 1340  
RESULT 7  
LOCUS AX664697 1245 bp DNA linear PAT 22-MAR-2003  
DEFINITION Sequence 6 from Patent WO02074960.  
ACCESSION AX664697  
VERSION AX664697.1 GI:29164457  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS 1 Leiby K.R., Kapeller-Libermann R. and Glucksmann M.  
TITLE Of human proteins and uses thereof  
JOURNAL Patent: WO 02074960-A 6 26-SEP-2002;  
Milleium Pharmaceuticals, Inc. (US)  
FEATURES  
source  
1. .1245  
/organism="Homo sapiens"

ORIGIN /mol\_type="unassigned DNA" /db\_xref="taxon:9606"

Alignment Scores:

Pred. No.: 9.1e-110 Length: 1245  
Score: 1203.00 Matches: 232  
Percent Similarity: 98.33% Conservative: 3  
Best Local Similarity: 97.07% Mismatches: 4  
Query Match: 97.02% Indels: 0  
Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AX664697 (1-1245)

QY 1 LeuLeuValaleuValaGlnGlyValaSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
Db CTGTGGCATATTAGTTCAAGAGAGTTTCAGATTTTCGCTTCTTGGAAGCTTTGGACATA 585  
QY 21 AlaProGlyAspMetLeuAspLysValaAlaArgArgLeuSerLeuIleLysHisProGlu 40  
Db GCACCAAGGTGACATGCTTGACAAAGGTGGCAAGAAACATTTCTTATATAAACATCCAGAG 645  
QY 41 CysSerThrMetSerGlyValaValaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
Db TGCTCCACCATGAGTGTGGGAAGCCATGAAACATTTGGCCAAACAAAGAAATGATTT 705  
QY 61 HisPheAspIleLysProProLeuHisIleAlaLysAsnCysAspPheSerPheThrGly 80  
Db CATTTTGACATCAAACTCCCTTGACATGCTAATAAATTGTGATTTTCTTTTACTGGA 765  
QY 81 LeuGlnHisValaThrAspLysIleIleMetLysGlyGluGlyGluGlyIleGluLys 100  
Db CTTCAACACGTTTACGATTAATAATGAAAGAAACAGAAACAAAGAAAGTATTGAGAAG 825  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValaGlnHisThrMetAlaCys 120  
Db GGGCAAAATCTGTCTTCAGACAGACATGCTGCACAGATACACACAAATGCGCATGT 885  
QY 121 HisLeuValaLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
Db CATTTGTGAAGAAGAACACATCGGGCTATTCGTTTGTGAACAGAGAACTGTTAACCT 945  
QY 141 GlnAsnAsnAlaValaLeuValaAlaSerGlyValaAlaSerAspPheThrIleArgArg 160  
Db CAATAATATGACATGCTGTTGCACTGCTGCTGCGCAAGTAATTCTATATCCGACGA 1005  
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180  
Db GCTCTGGAATTTTAAACAAACGCAACAGTGCATTTGTGTCTCTCCACAGACTA 1065  
QY 181 CysThrAspAsnGlyIleMetIleAlaThrPheAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
Db TGCACTGATTAATGGCATTATGATTCATGAAATGATTAAGAAAGCTACGTCCTGCTTG 1125  
QY 201 GlyIleLeuHisAspIleGluGlyIleArgArgIleArgProLysCysProLeuGlyValaAsp 220  
Db GGCAATTTTACATGACATAGAGAGCATCCGCTATGAAACAAATGTCTCTTGGATGAGC 1185  
QY 221 IleSerLysGluValaGlyGluAlaSerIleLysValaProGlnLeuLysMetGluIle 239  
Db ATATCAAAAGAAAGTTGGAGAAGCTTCATTAATAAGTACCAATTAATAATGAGATTA 1242

RESULT 8  
AX664695 1820 bp DNA linear PART 22-MAR-2003  
LOCUS AX664695  
DEFINITION Sequence 4 from Patent WO02074960.  
ACCESSION AX664695  
VERSION AX664695.1 GI:29164455  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1  
AUTHORS Leiby,K.R., Kapeller-Liebermann,R. and Glucksmann,M.  
TITLE 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions  
JOURNAL of human proteins and uses thereof  
Patent: WO 02074960-A 4 26-SEP-2002;  
Millennium Pharmaceuticals, Inc. (US)  
location/Qualifiers

FEATURES

source 1..1820  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

CDS

146..1390  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD80044.1"  
/db\_xref="GI:29164456"  
translation="MLILTKTGVFPFKSKRVPELSEFNHPGTLFLAKIVGIET  
SCDDPRAAVNDGTGNVLEAHSIQEVNHLKNGIYIPAAOOLHRENIQIYVEALSAS  
GVSPEDLSAIVATTIKPGLALSLGVLSFSLVGLVGLKRPFTIHMEBALITRLTNK  
VEPPLVAILISGGHCLALVQGVSDFLVGLSLDIPAGMDLKVARRLSLIRPECT  
MSGKALIEHLAKQGRFHPDIDKPLHAKNCDPSTGLOHVDKNNENKOEIEKG  
OILSSADIAIVQHTMACILVKTTHRAILPCKORDLIPONNAVVASGVASNFYIR  
RALRTLTATCTTLCPPPRCTDNGIMIANMGIIRLRAGLSGLIHDIISIRPEPKPL  
GVDISKEVSEASIKVPOLKMEI"

ORIGIN

Alignment Scores:

Pred. No.: 1.43e-109 Length: 1820  
Score: 1203.00 Matches: 232  
Percent Similarity: 98.33% Conservative: 3  
Best Local Similarity: 97.07% Mismatches: 4  
Query Match: 97.02% Indels: 0  
Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AX664695 (1-1820)

QY 1 LeuLeuValaleuValaGlnGlyValaSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
Db CTGTGGCATATTAGTTCAAGAGAGTTTCAGATTTTCGCTTCTTGGAAGCTTTGGACATA 730  
QY 21 AlaProGlyAspMetLeuAspLysValaAlaArgArgLeuSerLeuIleLysHisProGlu 40  
Db GCACCAAGGTGACATGCTTGACAAAGGTGGCAAGAAACATTTCTTATATAAACATCCAGAG 790  
QY 41 CysSerThrMetSerGlyValaValaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
Db TGCTCCACCATGAGTGTGGGAAGCCATGAAACATTTGGCCAAACAAAGAAATGATTT 850  
QY 61 HisPheAspIleLysProProLeuHisIleAlaLysAsnCysAspPheSerPheThrGly 80  
Db CATTTTGACATCAAACTCCCTTGACATGCTAATAAATTGTGATTTTCTTTTACTGGA 910  
QY 81 LeuGlnHisValaThrAspLysIleIleMetLysGlyGluGlyGluGlyIleGluLys 100  
Db CTTCAACACGTTTACGATTAATAATGAAAGAAACAGAAACAAAGAAAGTATTGAGAAG 970  
QY 911 CTTCAACACGTTTACGATTAATAATGAAAGAAACAGAAACAAAGAAAGTATTGAGAAG 970  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValaGlnHisThrMetAlaCys 120  
Db GGGCAAAATCTGTCTTCAGACAGACATGCTGCACAGATACACACAAATGCGCATGT 1030  
QY 971 GGGCAAAATCTGTCTTCAGACAGACATGCTGCACAGATACACACAAATGCGCATGT 1030  
QY 121 HisLeuValaLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
Db CATCTTGTAATAAAGAACACATCGGGCTATTCGTTTGTATAGAGAGAGACTGTTTACCT 1090  
QY 141 GlnAsnAsnAlaValaLeuValaAlaSerGlyValaAlaSerAspPheThrIleArgArg 160  
Db CAATAATATGACATGCTGTTGCACTGCTGCTGCGCAAGTAATTCTATATCCGACGA 1150  
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180  
Db GCTCTGGAATTTTAAACAAACGCAACAGTGCATTTGTGTCTCTCTCCACAGACTA 1210  
QY 181 CysThrAspAsnGlyIleMetIleAlaThrPheAsnGlyIleGluArgLeuArgAlaGlyLeu 200





Qy	181	CysThrAspArgnGlyTlIleMetTlIleAlaTTPanGlyYlIleGluArgLeuAlaGlyLeu	200
Db	650	TGCATGACCAATGGCATCATGATTCATGAGAAATGGAATGAAAGATACGTCCTGGCTTG	709
Qy	201	GlyIleLeuHlIleAspDlIleGluGlyIleAArgTlYArgIProLysCySPProLeuGlyValAsp	220
Db	710	GGCGTTTACATGATGATAGAGACATCCGATATGAACCAAAATGTCCTTGGAGTAGAC	769
Qy	221	IleSerIysGluValIGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle	239
Db	770	ATATCCAGAGAAGTGCAGAAAGCTCCCATPAAAAGTACCGCATTTAAAATGGCACTT	826
RESULT 11			
AX713716		2208 bp	linear
LOCUS			
DEFINITION	Sequence 400 from Patent EP1293569.		
ACCESSION	AX713716		
VERSION	AX713716.1	GI:29888642	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE			
AUTHORS	1 Isegai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R., Tanemichi,I., Seki,N., Yoshikawa,T., Otsuka,M., Negahari,K. and Masuho,Y.		
TITLE	Full-length cDNAs		
JOURNAL	Patent: EP 1293569-A 400 19-MAR-2003;		
FEATURES	Helix Research Institute (JP) ; Research Association for Biotechnology (JP)		
source	location/Qualifiers		
	1..2208		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Alignment Scores:			
Pred. No.:	3.57e-95	Length:	2208
Score:	1059.00	Matches:	211
Percent Similarity:	88.28%	Conservative:	0
Best Local Similarity:	88.28%	Mismatches:	4
Query Match:	85.40%	Indels:	24
DB:	6	Gaps:	1
US-10-649-273-2_COPY_176_414 (1-239) x AX713716 (1-2208)			
Qy	1	LeuLeuAlaLeuValIGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle	20
Db	869	CTGTGGCATTAAGTTCAAGAGATTTTCAATTTTCTGCTTGGAAACCTTTGGACATA	928
Qy	21	AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProLys	40
Db	929	GCACGAGGTGACATGCTTGACACAGGTGSCAAGAAACCTCTTAATAAAACATCCAAAG	988
Qy	41	CysSerTrmMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe	60
Db	989	TGCTTCACCATGATGATGCTGGAGAAACCATAGACATTTGGCCAAACAGAAATGATTT	1048
Qy	61	HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly	80
Db	1049	CATTTCAGATCAAACTCCCTTCGATCATGCTAATAAATTGTGATTTCTTTTACGGA	1108
Qy	81	LeuGlnHisValIThrAspLysIleIleMetLysLysGlyLysGluGluGlyTlIleGluLys	100
Db	1109	CTTCAACACGTTACGATGAATAATATATGAAAAAGAAAAAGGAAGGATTTGAGAAAG	1168
Qy	101	GlyGlnIleLeuSerSerSerHisHisAspIleHisAlaIleThrValGlnHisThrMetAlaLys	120
Db	1169	GGGCAATCCCTGCTTTCAGACAGACACATTCCTCCACAGTACAGCACACATGGCATGT	1228

[illegible]

TITLE	Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuna, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuh, Y., Nagai, K. and Isegai, T.
JOURNAL	NBDO human cDNA sequencing project
REFERENCE	Unpublished
AUTHORS	3 (bases 1 to 2208)
TITLE	Isegai, T., Otsuki, T. and Sugiyama, T.
JOURNAL	Direct Submission
	Submitted (24-OCT-2001) Takao Isegai, Helix Research Institute,

**COMMENT**

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

```

FEATURES
source
location/Qualifiers
1..2208
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="FEBRA2004592"
/tissue_type="brain"
/clone_id="FEBRA2"
/dev_stage="fetus"
/note="Cloning vector: pME18SFL3"
344..1438
CDS

```

```

/note="unnamed protein product"
/codon_start=1
/protein_id="BAB70923.1"
/db_xref="GI:16550167"
/translation="MLILITKTAAGVFFKPSKKRVEFLRSFNFHPGTLFLAKHIVLAGIET
SCDDTAADVDEGTGALVEALHSQTEVHLKTCGAIYPPAAQOLHRENIORIVQALASAS
GVSQSDLSAIVTTIKPGALSLGVLSLQVQLQKRPFIPIHMEAHALITRLTNL
VEPFFIVLILSGGHCLALAVQGVSDFLKNGSDIAPDMDLKVARRPLIKHPES
MGGKAIYHLAKQGNRRFFPDIKPLAHNAKNDSPFTQYQHTYDKIIMKGEABEGEKK
QILSSADIATAVQHTMACHLVKTRTHAHLFCKQDILLPÖNNAAVAVASGVAASFCIL
RALRILTNATQCTLLCPPRLCTDNGIMIA"

```

Alignment Scores:	
Pred. No.:	3,576-95
Score:	1059.00
Percent Similarity:	88.28%
Best Local Similarity:	88.28%
Query Match:	85.40%
GB:	9
	Gaps: 1
	Length: 2208
	Conservative: 211
	Mismatches: 0
	Indels: 24

US-10-649-273-2\_COPY\_176\_414 (1-239) X AK055441 (1-2208)

Qy	1	leu	en	u	a	l	a	e	v	a	l	a	l	g	n	g	l	y	v	a	s	e	r	a	s	p	h	e	l	e	u	e	l	e	u	g	l	y	l	y	s	e	r	e	n	u	s	p	i	e	20					
Dd	869	c	t	t	g	t	t	g	c	a	t	t	a	g	c	a	a	g	a	t	t	c	a	g	a	t	t	t	c	t	t	c	t	t	g	a	a	a	g	c	t	t	t	g	a	c	a	n	a	928						
Qy	21	a	l	a	a	p	r	o	g	l	y	a	s	p	h	e	l	e	u	s	p	i	y	s	a	l	a	a	r	g	a	r	g	l	e	u	s	e	r	l	e	u	i	l	e	y	s	i	s	p	r	o	g	i	u	40
Dd	929	g	c	a	c	a	c	a	g	a	t	t	c	a	a	g	t	t	g	a	c	a	a	g	t	t	c	c	t	t	t	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	988						
Qy	41	c	y	s	e	r	t	h	i	m	e	s	e	r	g	i	y	l	y	s	a	l	a	l	e	g	i	u	i	s	e	u	a	l	a	l	y	s	g	i	n	g	i	y	s	n	a	r	p	h	e	60				
Dd	989	t	g	c	t	c	c	a	c	a	t	t	a	g	a	t	t	g	t	g	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	1048							
Qy	61	h	i	s	p	h	e	a	s	p	i	e	l	e	y	u	s	p	r	o	t	e	u	i	s	i	s	a	l	a	l	y	s	a	s	e	r	a	s	p	h	e	s	e	r	p	h	e	r	g	i	u	80			
Dd	1049	c	a	t	t	t	t	t	g	a	c	a	t	t	c	c	t	t	g	a	c	a	r	e	c	t	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	1100						
Qy	81	l	e	u	g	i	n	h	i	s	v	a	l	t	h	r	a	s	p	i	s	i	l	e	n	e	t	y	s	l	y	s	g	i	u	s	g	i	n	g	i	y	l	e	g	i	u	s	100							

Db	1109	CTTCGACACGCTTACTGTAAATAATATATGACAAAAAGAAAAAGGAGAGATATTGAGAG	1168
Oy	101	GIYGLINILEUSTRSERIALAALASPIILEALALATHRALGINHSTRMETALA	Cys 120
Db	1169	GGGCAAAATCCGTCTTGACAGCAGACATTCCTGCCAGATACGACACAAATGGCAGT	1228
Oy	121	HISLEUVALLYSARGTHRHSARGALALEUENPHECYSLVSGINAGSAPLEUENPRO	140
Db	1229	CATCTTGTGAAAAACACATCGGGCTATTCTGTTTGTAAAGCAGACGCTGTACT	1288
Oy	141	GLNASHNAPLAVALLLEUVALALASERGLVVALALASERANPHELYRILEARG	160
Db	1289	CAAAATATATGACATACGGTGTGACATCTGGTGSGTCCGAAGTAACTTCTGATCCGACGA	1348
Oy	161	ALALEUGLILEUSTRASNAALATHGLCYSTRILEUENCYSPROPROARGLEU	180
Db	1349	GCTCTGGAATTTTAAACAAACGACACACAGGACCTTGTGTGTCCTCCACGACTA	1408
Oy	181	CYSTRASPAENGLYILEMETILEALATRPASNGLYLEUGARGLEUARGALAGLYLEU	200
Db	1409	TGCCTGATTAATGGCATTATGATTGCA-----	1435
Oy	201	GLYILEUHSASPILEGUGLYLEARGTYRGLUPROLYSCYSPROLEUGLYVALASP	220
Db	1436	-----TGATGTCTCTTTGAGAGTAC	1456
Oy	221	ILESERYSGLUVALGLYGLUVALASERILEYSLVALPROGLEUENLYSMETGLNILE	239
Db	1457	ATATCAAAAGAGATGGAGAGACTTCATTAATAATACACACATTTAAATAGAGATA	1513

RESULT	13
LOCUS	BC078974
DEFINITION	BC078974 1546 bp mRNA linear ROD 03-AUG-2004
ACCESSION	Rattus norvegicus cDNA clone IMAGE:7111906, partial cds.
VERSION	BC078974
KEYWORDS	BC078974.1 GI:50926879
SOURCE	.
ORGANISM	Rattus norvegicus (Norway rat)
	Rattus norvegicus

JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED	12477922
REFERENCE	2 (bases 1 to 1546)
AUTHORS	Director MGC Project.
TITLE	Direct Submission
JOURNAL	Submitted (02-AUG-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>



COMMENT

Contact: MGC help desk  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Howard Jacobs  
cDNA Library Preparation: Express Genomics  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) mcd@paxil.stanford.edu  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Series: IRAX Plate: 182 Row: f Column: 6  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.  
Location/Qualifiers

FEATURES

Source

1. 1546  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
/clone="IMAGE:7111906"  
/tissue\_type="Testis, rat (Brown Norway)"  
/clone\_lib="NIH MGC\_237"  
/lab\_host="DH10B"  
/note="Vector: pExpress1"  
144. >1546  
/codon\_start=1  
/product="unknown (protein for IMAGE:7111906)"  
/protein\_id="AAH78974.1"  
/db\_xref="GI:50926880"  
/translation="MTLSTKATGATPPRPSNVGRIRBNVOPALFHKIVGIEIT  
SCDTPAAVDETVGALGSAHSQTEVHLKGTGTPPAQGLHRENIQRIVEALSSAS  
GSPSDLSAIIATIKPGLALSDVGLSVSVQVNFKKPFIPIHMEALATIRLTHK  
VGPPELVLLISGHCFLALVQSVDFLLGSLDIPDMDKVARRLSLIGPEST  
MSGKAIHLAKENGRFHTINPWNACNCFSTGLQHTVDKLTHERKEGIEKG  
OILSSADIAAIVOHATACHLAKTRHAILFCQONLSPAAVLVSGVASENLYIR  
RALEIVANATCTLCPPRLCTDNGIMIANNGIRLALAGCILDPVEDIRERPAGE  
IVSMRVITATTELDSDSVSHHTVAHSPINSSGRANIQTSKCTQTYVMKTRHIL  
NINLRSKKKKKKKKK"

CDS

Alignment Scores:

Pred. No.: 8.27e-88 Length: 1546  
Score: 983.00 Matches: 188  
Percent Similarity: 92.06% Conservative: 9  
Best Local Similarity: 87.85% Mismatches: 17  
Query Match: 79.27% Indels: 0  
DB: 10 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x BC078974 (1-1546)

Qy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIle 20  
Db 669 CTGTTGGCGTTAGTTCAAGAGTCTTTCAGATTTCTGCTCTCGGAGAGCTCCGACATTA 728  
Qy 21 AlaProGlyAspMetLeuAspIleValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
Db 729 GCGCAGGCGCATCTTGACAGAGTGCGCAAGAACATTTCTTTATATCAATCATCAGAA 788  
Qy 41 CysSerThrMetSerGlyGlyLeuAlaIleGluLysLeuAlaLysGlnGlyAsnArgPhe 60  
Db 789 TGTTCCTCAATGAGTGGGAGAAAGCTATAGAACATTTGGCCCAAGAAAGAAATAGATTTC 848  
Qy HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
Db 849 CACTTTACTATCAATCAACCCATGCAAGATGTAAGAACTGATTTTCTTTACGGGGA 908  
Qy LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100  
Db 909 CTTCAACATGTCAACCGATTAAGCTAATTAACACACAGAAAGAAAGAGCATTTAGAGAG 968

ORIGIN

Qy 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120  
Db 969 GGGCAATCTCTGTCAATCAGCCGACATTCGCTGCGGTACAGACGCAACAGCGCTGC 1028  
Qy 121 HisLeuValLysArgThrHisArgAlaAlaLeuPheCysLysGlnArgAspLeuPro 140  
Db 1029 CACTTCGGAAAGAACACATTCGTATTCCTTTTGGCCAGAGAAATTTGCTATCTT 1088  
Qy 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArg 160  
Db 1089 CCAGCTAACCGCATATTAGTTGTCTGTGAGAGTGTTCAGAGTAATCTGTACATCCAGA 1148  
Qy 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
Db 1149 GCATTGGAATTTGTACAAATGCAACACATGCAATGCAATGCTTTGTTGTCCCTCCGAG 1208  
Qy 181 CysThrAspAsnGlyIleMetIleAlaIleArgGlyIleGluArgLeuAlaGlyLeu 200  
Db 1209 TGCACCTGACAAATGCTATCATGATTGCAATGGAATGAATGAATTAACGTGCTGCTTG 1268  
Qy 201 GlyIleLeuHisAspIleGluGlyIleArgTyrlleGluProLys 214  
Db 1269 GGCATTTTACATGATGATGAGACATCCGATACGAACCAAG 1310  
RESULT 14  
BC051211 1109 bp mRNA linear ROD 15-APR-2003  
LOCUS Mus musculus, clone IMAGE:1327545, mRNA.  
DEFINITION BC051211  
VERSION BC051211.1 GI:29881634  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM  
Tissue: Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1109)  
Strausberg, R.  
Direct Submission  
Submitted (14-APR-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Marcello Bento Soares, Ph.D.  
cDNA Library Preparation: M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
contact: amadan@systemsbiology.org  
Anup Madan, Jeessia Fahey, Erin Helton, Mark Kettelman, Anuradha  
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

REMARK

COMMENT

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Series: IRAX Plate: 113 Row: b Column: 1  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.  
Location/Qualifiers

FEATURES

Source

1. 1109  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1327545"  
/tissue\_type="Thymus gland, mouse"  
/clone\_lib="Soares\_chymus\_2NBMT"  
/lab\_host="DH10B"  
/note="Vector: p773D-Pac"

ORIGIN





```

Qy 161 AlaleuGlulIleuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
Db 574 GAGCTGCAGACTCTGGCAAAATGCAAACGGTTTGCTTTCTGTCTCTCTCTCAAGGCTG 633
Qy 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 634 TGCACCGATATATGGGTATATGATGCAATGCATGAAATGCATTGAAAGGTGCGTCAGAGATGT 693
Qy 201 GlyIleLeuHisAspIleGluGlyIleArgTyrgIupProLysCysAsProLeuGlyValAsp 220
Db 694 GGATTTTATACAGTACTGATGGCATCCGCTACGAACCAAAAGCTCCCTTGGAATTGAT 753
Qy 221 IleSerIysGluValGlyGluAlaSerIleLysValProGlnLeuLys 236
Db 754 ATTTCCAAAAGATTGAAGAGATTCCATCAAAAGTCCCAAGACTAAG 801

```

Search completed: November 10, 2005, 22:36:56  
 Job time : 4721.79 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - nucleic search, using frame\_plus.p2n model

Run on: November 10, 2005, 16:48:13 ; Search time 538.789 Seconds  
(without alignments)  
2625.922 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 LIALVQGVSDFLILGKSLDI.....DISKVGESAIKVPOLKMEI 239

Scoring table:

BLAST62	Xgapop 10.0 ,	Xgapext 0.5
Ygapop 10.0 ,	Ygapext 0.5	
Fgapop 6.0 ,	Fgapext 7.0	
Delop 6.0 ,	Delext 7.0	

Searched: 4390206 segs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DRV=x1D  
-O=/cgn2\_1/USFTO\_spool\_p/US10649273/runat\_02112005\_091337\_15540/app\_query.fasta\_1.1429  
-DB=N.Geneseq\_16Dec04 -QPM=faetap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=b1ts -START=1 -END=-1 -MATRIX=blslosum62 -TRANS=human40.cdi  
-LIST=45 -DOCLIN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=p2to -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10649273 @CGN 1.1 1063 @runat\_02112005\_091337\_15540 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : N.Geneseq\_16Dec04:\*  
1: geneseqn19808:\*  
2: geneseqn19908:\*  
3: geneseqn20008:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1240	100.0	1416	8	ABX70950
2	1240	100.0	1526	6	ABX70950 Novel hum
3	1240	100.0	2058	6	ABX70950 DNA encod
4	1240	100.0	2197	6	ABX70950 Human O-s
5	1240	100.0	2572	8	ABX70950 DNA encod
					ABX70950 Human pro

6	1203	97.0	1820	6	AA046856	AA046856 Human gly
7	1203	97.0	1820	8	ACA60887	ACA60887 Human cDN
8	1203	97.0	1821	10	ABX57020	ABX57020 cDNA enco
9	1059	85.4	2208	10	ADA52832	ADA52832 Human cod
10	1059	85.4	2890	12	AD024627	AD024627 Human sof
11	995.5	80.3	3358	10	AD031345	AD031345 Human dia
12	870	70.2	1572	6	AB075508	AB075508 Murine s1
13	599	48.3	2734	5	AA846422	AA846422 DNA encod
14	468	37.7	371	12	ADL86725	ADL86725 DNA up-re
15	468	37.7	371	12	ADL86726	ADL86726 DNA up-re
16	332.5	26.8	1601	4	AB124633	AB124633 Drosophi
17	332.5	26.8	3656	4	AB124632	AB124632 Drosophi
18	332	26.0	1557	3	ACA38454	ACA38454 Arabidops
19	257	20.7	1146	8	ACA26804	ACA26804 Prokaryot
20	248.5	20.0	1000	6	AAE91424	AAE91424 Moraxella
21	248.5	20.0	1000	6	ABX37804	ABX37804 DNA seque
22	248.5	20.0	1044	8	ACA39102	ACA39102 Prokaryot
23	248.5	20.0	94750	4	AAE28551	AAE28551 Genomic E
24	247	19.9	936	8	ACA20445	ACA20445 Prokaryot
25	247	19.9	1053	12	ADL03120	ADL03120 DNA encod
26	241	19.4	1032	8	ACA43173	ACA43173 Prokaryot
27	239	19.3	4360	6	AA048239	AA048239 Ehrlichia
28	235.5	19.0	1092	6	AB090383	AB090383 M. caprai
29	234	18.9	300	3	AAH00934	AAH00934 Human cDN
30	232	18.7	676	4	AAH08019	AAH08019 Human cDN
31	232	18.7	1385	4	AAH15110	AAH15110 Pseudomon
32	231	18.6	1026	4	AA554064	AA554064 Pseudomon
33	231	18.6	1026	6	ACA42146	ACA42146 Prokaryot
34	231	18.6	1026	10	ADG73341	ADG73341 P aerugin
35	231	18.6	1059	11	ABD02280	ABD02280 Pseudomon
36	231	18.6	1206	11	ABD02197	ABD02197 Pseudomon
37	229	18.5	1026	10	ADG73343	ADG73343 P aerugin
38	224	18.1	1029	4	AA553109	AA553109 Haemophil
39	224	18.1	1029	8	ACA43150	ACA43150 Prokaryot
40	224	18.1	11000	2	AA742063	AA742063 Continnation (6 of
41	217	17.5	9567	13	ADT05493	ADT05493 Haemophil
42	217	17.5	85814	13	ADT05644	ADT05644 Haemophil
43	214	17.3	1044	4	AB108591	AB108591 Drosophi
44	213	17.2	1014	4	AA556045	AA556045 Salmonell
45	213	17.2	1014	8	ACA51431	ACA51431 Prokaryot

## ALIGNMENTS

RESULT 1	ABX70950	standard; cDNA; 1416 BP.
ID	ABX70950	standard; cDNA; 1416 BP.
XX	ABX70950;	
AC	ABX70950;	
XX	ABX70950;	
DT	05-MAR-2003	(first entry)
XX	05-MAR-2003	(first entry)
DE	Novel human cDNA sequence #175.	
XX	Novel human cDNA sequence #175.	
KW	Human; gene; ss; nervous system disorder; peripheral neuropathy;	
KW	Huntington's disease; amyotrophic lateral sclerosis; haemophilia;	
KW	neurodegenerative disease; Parkinson's disease; Alzheimer's disease;	
KW	autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;	
KW	insulin-dependent diabetes mellitus; anaemia; thrombocytopenia; wound;	
KW	ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;	
KW	fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;	
KW	coagulation disorder; cancer; tumour; inflammatory disease; septic shock;	
KW	Crohn's disease; anaphylaxis; proliferation; chemotactic;	
KW	differentiation; stem cell growth factor; hematopoiesis; chemokine;	
KW	haemostatic; antiinflammatory; expressed sequence tag; EST.	
OS	Homo sapiens.	
XX	Homo sapiens.	
PN	WO200281731-A2.	
XX	WO200281731-A2.	
PD	17-OCT-2002.	
XX	17-OCT-2002.	
PF	29-JAN-2002; 2002WO-US001222.	

XX 30-JAN-2001; 2001US-00774528.  
XX (HYSE-) HYSEO INC.  
XX (GOOD/) GOODRICH R. W.  
XX Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F,  
XX Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;  
XX WPI; 2003-058563/05.  
XX  
XX Novel polypeptide useful for treating neurodegenerative diseases, myeloid  
XX or lymphoid cell disorders, bone disorders, mechanical and traumatic  
XX disorders, coagulation disorders, and inflammatory diseases.  
XX  
XX Claim 1; Page; 612pp; English.  
XX  
XX This invention relates to the cDNA sequences encoding an isolated novel  
XX human polypeptide. The protein encoded by the nucleic acid of the  
XX invention is useful for treating central and peripheral nervous system  
XX diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic  
XX lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,  
XX Alzheimer's disease); autoimmune disease (e.g. systemic lupus  
XX erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)  
XX ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopenia)  
XX ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,  
XX osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head  
XX trauma); lung or liver fibrosis; reperfusion injury in various tissues;  
XX bacterial, viral or fungal infections; allergic conditions such as  
XX allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);  
XX cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's  
XX disease, anaphylaxis). The protein may be used to inhibit the growth,  
XX infection or function of infectious agents such as bacteria, fungi,  
XX viruses, or to effect bodily characteristics, biorhythms or circadian  
XX cycles of rhythms. The protein may also have  
XX proliferation/differentiation, stem cell growth factor, haematopoietic  
XX regulation, immune stimulating or suppressing, chemocactive/chemokinetic,  
XX haemostatic and thrombolytic, receptor/ligand, and antiinflammatory  
XX activities. The cDNA sequences of the invention are useful for expressing  
XX recombinant protein for analysis. The present sequence represents a novel  
XX human cDNA sequence of the invention, this sequence is an expressed  
XX sequence tag (EST) and was identified using subtractive hybridisation  
XX  
XX  
SQ Sequence 1416 BP; 441 A; 280 C; 273 G; 422 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 3,21e-132 Length: 1416  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0  
US-10-649-273-2\_copy\_176\_414 (1-239) x ABX70950 (1-1416)  
QY 1 LeuLennalaLeuValGInglYValSerAspHeLeuLeuGluGlySerLeuAspIle 20  
DB 586 CTGTGGCAATTAAGTTCAGAGAGGTTTCAGATTTTCGCTTTCGAAAGTCTTTGGACATA 645  
QY 21 AlaProGlyAspMetLeuAspLysValAlaAlaArgArgLeuSerLeuIleLysHisProGlu 40  
DB 646 GCACGAGGTGACATGCTTGACAGAGGTGCGCAAGAGACTTTCTTTAATAAATCATCAAG 705  
QY 41 CysSerThrMetSerGlyGlyLysAlaAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
DB 706 TGCTCCACCATGAGTGGTGGGAAAGCCATGAACAATTGGCCAAACAGAAATAGATT 765  
QY 61 HisPheAspLysLysProProLeuHisHisAlaLysAsnCyGAspPheSerPheThrGly 80  
DB 766 CATTTGACATCAAAACCTCCCTTCATCATCTAATAAATTTGATTTTCTTTAATCTGA 825  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100

DB 826 CTTCAACAGCTTATCGATTAATAATATGAAAAAGAAAAAGACAGACTATTGGAAG 885  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120  
DB 886 GGGCAAAATCTCTCTTCAGACAGACATGTGTCACAGATGACAGACAAATGGATGT 945  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
DB 946 CATCTTGAAAAAGAACATCGGGCTATTCTGTTTGAACAGAGACATGTTACCT 1005  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheTrpIleArgArg 160  
DB 1006 CAAATTAATGCGATGCTGTTGCATCTGTGTGTGCGCAAGTAATCTTATATCCGACA 1065  
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGlyLeu 180  
DB 1066 GCTCTGGAATTTTAACAAAGCCACACAGTCATCTTGTGTCTCTCCCAACAT 1125  
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
DB 1126 TGCACTGATTAATGCAATTATGATTCATGGAATGGATTAAGAACTACGTCGCTGG 1185  
QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyTrpGluProLysCysProLeuGlyValAsp 220  
DB 1186 GGCATTTTAACATGACATAGAAAGCATCCGCTATGAACCAAAATGTCCTCTTGAGATGAC 1245  
QY 221 IleSerLysGluValGlyGluAlaSerIleLysValAlaProGluLeuLysMetGluIle 239  
DB 1246 ATATCAAAAGAGTTTGAGAGAGCTTCATTAAGTACCAATTAAATGGAGATA 1302  
RESULT 2  
ABST76639 ID  
ID ABST76639 standard; DNA; 1526 BP.  
XX  
XX ABST76639;  
XX  
XX 11-DEC-2002 (first entry)  
XX  
XX DNA encoding novel human metalloprotease MP1 fragment #1.  
XX  
XX  
XX Metalloprotease, MP-1; immune disorder; glutamate transport; cancer;  
XX motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
XX reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
XX genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
XX Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
XX Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
XX liver disease; renal disease; immune disorder; rheumatoid arthritis;  
XX acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
XX emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
XX neurological disorder; gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200272751-A2.  
XX  
XX  
XX 19-SEP-2002.  
XX  
XX 05-FEB-2002; 2002WO-US003353.  
XX  
XX  
XX 05-FEB-2001; 2001US-0266518P.  
XX PR 10-APR-2001; 2001US-0282814P.  
XX  
XX  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
XX  
XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S,  
XX  
XX P-PSDB; ABG96487.  
XX  
XX WPI; 2002-723329/78.  
XX  
XX  
XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
XX treating, or ameliorating diseases associated with aberrant  
XX metalloproteinase activity, e.g. immune, metabolic, inflammatory and  
XX neurological disorders.  
XX  
XX  
XX

XX Disclosure; Page 462-463; 473pp; English.  
XX  
XX The invention describes an isolated nucleic acid molecule (I) encoding a  
CC metalloprotease (MP-1). (I) is useful for preventing, treating, or  
CC ameliorating a medical condition, particularly an immune disorder, an  
CC aberrant glutamate transport or motor neuron disorder, such as an  
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
CC condition. The compositions and methods are also useful for diagnosing,  
CC prognosticating, treating, ameliorating and/or treating disorders  
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease). This sequence represents a  
CC metalloprotease MPI polynucleotide  
XX

Sequence 1526 BP; 484 A; 297 C; 289 G; 456 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	3,57e-132	Length:	1526
Score:	1240.00	Matches:	239
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2\_COPY\_176\_414 (1-239) x ABA93268 (1-1526)

QY 1 LeuenualaValGlnGlyValSerapheLeuenuGlyVysSerleuaple 20  
DB 85 CTGTTGCATTAGTTCAGAGAGGTTTCAGATTTCGCTTTCGAAAGCTTTGACATTA 144  
QY 21 AAlaproglyaspmetleuaspVAlAaRgRgLeuSerleuileVysHisProglu 40  
DB 145 GCACCAAGTGCACGCTTGCACAGGTGCAGAAAGACTTTCTTAAATAAACATCAGAG 204  
QY 41 CysSerThermetSerglyVlyVAlaAlleGlnHisleuAlaVysGlnGlyVAsnaRgpe 60  
DB 205 TGCTCCACATGATGTGGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATGATT 264  
QY 61 HispheapRlelyPProProleuHisAlaVysaenCyEapRphSerPheThrgly 80  
DB 265 CATTTCACATCAAAACCTCCCTTCATCATCTAAAAATTTGATTTTCTTTACTGGA 324  
QY 81 LeuGlnHisValThrApRlyVlellemetVysVysGlnVysGlnGlyVlelleGluys 100  
DB 325 CTTCACACGCTTACTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 384  
QY 101 GlyGlnleuSerSerAlaAlaAspRleAlaAlaThrValGlnHisThermetAlaCys 120  
DB 385 GGGGAATTCCTGCTTGCAGCAGACATTGCTGCACAGTACAGACACAAATGCGATGT 444  
QY 121 HisleuVallyVaygThrHisArgAlaAlleuPheCyVysGlnVysGlnVysAspRleuPro 140  
DB 445 CATCTTGAAGAAACACATCGGGCTATTCGTTTGTAAAGACAGAGACTTTTACTCT 504  
QY 141 GlAsnaRnaAlaValleuValAlaSerGlyVlyVAlaAlaSerapheThrglyleargarg 160  
DB 505 CAATAATATGACATGCTGTTGCATCGTGCTGTCGCAAGTAATCTTATATCCGCGAGA 564  
QY 161 AAlaenuGluileuThrAsnaAlaThrglnCysThrleuenuCysPProProRargleu 180  
DB 565 GCTCTGGAATTTTAACAAACGACACAGTCACTTTGTTGTGTCTCTCCCGACATTA 624  
QY 191 CysThrAsnaRnglyVlelleVlelleVlelleVlelleVlelleVlelleVlelleVlelle 200  
DB 625 TGCACTATATATGACATTAATGATTCAGATGAAATGATTAATTAATTAATTAATTAAT 684

QY 201 GlylleuenuHisApRlelleGlnGlyVlelleArgTgRgluProlyCysPProleuGlyVAlaAp 220  
DB 685 GGCATTTTACATGACATAGAGGATCCGCTATGAACAAAATGTCCTCTTGAGATGAC 744  
QY 221 HisSerlyGlnVAlGlyGlnVAlaSerlelyVAlProglleuVysmetGluile 239  
DB 745 ATATCAAAAGAAAGTTGAGAGAGCTTCATTAAGTACCAAAATTAATTAATTAATTAAT 801

## RESULT 3

ABA93268  
ID ABA93268 standard; cDNA; 2058 BP.

AC ABA93268;

DT 19-APR-2002 (first entry)

DE Human O-sialoglycoproteinase-like protein encoding cDNA SEQ ID NO:1.

XX Human; O-sialoglycoproteinase-like protein; OSGPLP; enzyme; gene; ss.

XX Homo sapiens.

OS Homo sapiens.

PH Key Location/Qualifiers

FT CDS 110..1354

FT /\*tag= a

FT /product= "O-sialoglycoproteinase-like protein"

PN CN1318550-A.

PD 24-OCT-2001.

PP 19-APR-2000; 2000CN-00106834.

PR 19-APR-2000; 2000CN-00106834.

PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.

PI Mao Y, Xie Y;

PT MPI: 2002-115090/16.

DR P-PSDB; ABB05481.

XX O-sialoglycoproteinase-like protein and encoding polynucleotide, useful for diagnosing, preventing and treating related diseases.

XX Claim 5; Page 29-30 (Disclosure); 38pp; Chinese.

XX The present sequence encodes human O-sialoglycoproteinase-like protein (OSGPLP). The present invention also describes: (1) the preparation of the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the prevention and/or treatment of related diseases; (4) utilizing the OSGPLP protein in screening its agonist, excitomotor and inhibitor and preparing an antibody against the OSGPLP protein; and (5) the use of the OSGPLP

XX CC polynucleotide sequences, proteins, agonists, excitomotors, inhibitors and antibodies in treating diseases related to the abnormal OSGPLP gene and in preparing the medicine composite for the treatment

XX SQ Sequence 2058 BP; 637 A; 400 C; 410 G; 611 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	5,45e-132	Length:	2058
Score:	1240.00	Matches:	239
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2\_COPY\_176\_414 (1-239) x ABA93268 (1-2058)

QY 1 LeuenualaValGlnGlyValSerapheLeuenuGlyVysSerleuaple 20  
DB 635 CTGTTGCATTAGTTCAGAGAGTTCAGATTTCGCTTTCGAAAGCTTTGACATTA 694

Qy	21	AlaProGlyAspMetLeuAspLysValAlaAspGluSerLeuLysHisProGlu	40
Db	695	GCACCAAGGTGACATCTCTTGCACAAAGGTGCAGAAAGACTTTCTTTATATAACATCCAGAG	754
Qy	41	CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe	60
Db	755	TGCTCCACCATAGAGTGTTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATATAGATTT	814
Qy	61	HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly	80
Db	815	CATTTTGCATCTCAACACTCTCCCTTCATCATTCATTAATAAAATTTGATTTTCTTTACTGGA	874
Qy	81	LeuGlnHisValThrAspLysIleIleMetCysLysGlnLysGlnGlyIleGluLys	100
Db	875	CTTCAACACGTTACTGATATAATATATATAAAGAAAGAAAGAGAGATTATGAGAG	934
Qy	101	GlyGlnIleLeuSerSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCys	120
Db	935	GGGCAAACTCTGCTCTTCCAGCACAGACATTGGCTGCCACAGTACAGACACAAATGGCATGT	994
Qy	121	HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro	140
Db	995	CATCTTGGAAGAAAGAACATCGGGCTATTCTGTTTGTATGACAGAGAGACTTGTTACCT	1054
Qy	141	GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspAspPheTrpIleArgArg	160
Db	1055	CAAAATATATGACGTCATCGTTCATCTGATCTGGTGGTGCAGATACCTTCATATATCCGACGA	1114
Qy	161	AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu	180
Db	1115	GCTCTGGAAATTTTAAACAAACCCGACACACAGTGCATCTTGTCCTCTCCACAGCTA	1174
Qy	181	CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu	200
Db	1175	TGCACTGATTAATGCGATTAATGATTCATGATGATGAATGATTAAGAACTACGTCGCGCTTG	1234
Qy	201	GlyIleLeuHisAspIleGlnGlyIleArgTrpGlnProLysCysProLeuGlyValAsp	220
Db	1235	GGCATTTTAAACATGACATAGACATAGAGCATCCGCTATGAAACAAAATGTCCTCTTGGAGTAGAC	1294
Qy	221	IleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGluIle	239
Db	1295	ATATCAAAAGAAAGTTGAGAGAGCTTCATATAAAGTACCAATATATAAATGAGAGATA	1351
RESULT 4			
ABST7635			
ID	ABST7635	standard; DNA; 2197 BP.	
XX	AC	ABST76635;	
XX	DT	11-DEC-2002 (first entry)	
XX	DE	DNA encoding novel human metalloproteinase MP1.	
XX	KM	Metalloproteinase, MP-1; immune disorder; glutamate transport; cancer;	
XX	KM	motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;	
XX	KM	reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;	
XX	KM	genital wart; metabolic disorder; premature puberty; Kallman syndrome;	
XX	KM	Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;	
XX	KM	Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;	
XX	KM	liver disease; renal disease; immune disorder; rheumatoid arthritis;	
XX	KM	acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;	
XX	KM	emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;	
XX	KM	neurological disorder; gene; ds.	
OS	XX	Homo sapiens.	
XX	PN	WO200272751-A2.	
XX	PD	19-SEP-2002.	
XX	PF	05-FEB-2002; 2002WO-US003353.	

[illegible]

DB 1116 CATCTTGTGAAAAAGAACACATCGGCTATTCTGTTTGTAGACAGAGACTTGTACT 1175  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaAspAsnPhenylleArgArg 160  
DB 1176 CAAATATATGACAGTACTGTTGATCTGGTGTCGCCAAGTAACTTCTATATCCGAGCA 1235  
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrIleuLeuCysProProProArgLeu 180  
DB 1236 GCTCTGAAATTTTAACTCAAAACGACACAGTGCCTTGTGTCTCTCCCAAGACTA 1295  
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluValLeuArgAlaGlyLeu 200  
DB 1296 TGCATGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1355  
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220  
DB 1356 GGCATTTTACATACATAGAGAGATCCGCTATGACCAAAATCTCTTGGAGTAC 1415  
QY 221 ILeSerLysGluValGlyValAspIleLysValProGluLeuLysMetGluIle 239  
DB 1416 ATATCAAAAGAAAGTGGAGAGCTTCATTAAGTACACAAATTAAGTACATA 1472  
RESULT 5  
ABT23207  
ID ABT23207 standard; DNA; 2572 BP.  
AC ABT23207;  
XX  
XX 01-MAY-2003 (first entry)  
XX  
DE Human protein modification + maintenance molecule DNA SEQ ID No 36.  
XX  
XX Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;  
KM cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;  
KM antituber; hepatotropic; gynecological; antibacterial; virucide;  
KM protozoecide; antiparasitic; cell proliferative disease; PMOD;  
KM protein modification and maintenance molecule; immunogenic fragment;  
KM cancer; autoimmune; inflammatory disease; neurological disorder;  
KM gastrointestinal; developmental; vesicle trafficking disorder; infection;  
KM protein-protein interaction; drug-target interaction;  
KM gene expression profile; human; gene; ds.  
OS Homo sapiens.  
XX  
XX MO2003000844-A2.  
PN 03-JAN-2003.  
XX  
XX 18-JUN-2002; 2002MO-US019360.  
PF  
XX 22-JUN-2001; 2001US-0300508P.  
PR 06-JUL-2001; 2001US-0303445P.  
PR 13-JUL-2001; 2001US-0305405P.  
PR 09-AUG-2001; 2001US-031442P.  
PR 24-AUG-2001; 2001US-0314821P.  
PR 29-AUG-2001; 2001US-031592P.  
PR 03-MAY-2002; 2002US-0378205P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
PA  
PI Gandhi AR, Kable AE, Swarnakar A, Hafajia AJA, Tran B, Duggan BM;  
PI Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DM, Lee EH, Yue H;  
PI Forsythe TJ, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J;  
PI Thangavelu K, Gietzen KJ, Ding L, Baughn MK, Borowsky ML, Yao MG;  
PI Walla MK, Mason PM, Gururajan R, Lee S, Becha SD, Lee SY, Tran UK;  
PI Elliott VS, Luo W, Sprague MW, Tang YT, Lu Y, Zebajadian Y;  
XX  
XX WPI: 2003-184039/18.  
DR P-PSDB; ABU26654.  
XX  
XX New isolated human PMOD polypeptide and polynucleotide, useful for  
PT diagnosing, treating and preventing diseases or conditions associated  
PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and

PT infections.  
XX  
XX Claim 91; Page 211; 225pp; English.  
XX  
XX The invention relates to an isolated polypeptide comprising: any of 28  
CC sequences of 48-1256 amino acids; a natural amino acid sequence at least  
CC 90% identical to the 28 amino acid sequences; 94% identical to a sequence  
CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino  
CC acids, or 97% identical to a sequence of 242 amino acids, all given in  
CC the specification; or a biologically active or immunogenic fragment of  
CC the isolated polypeptide. The polypeptides and polynucleotides are useful  
CC in diagnosing, treating and preventing diseases or conditions associated  
CC with the decreased expression of protein modification and maintenance  
CC molecules (PMOD), such as cell proliferative diseases (e.g. cancer,  
CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,  
CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,  
CC epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.  
CC endometriosis), developmental, vesicle trafficking disorders, and  
CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also  
CC useful in assessing the effects of exogenous compounds on the expression  
CC of nucleic acid and amino acid sequences of PMOD. The PMOD or its  
CC fragments are useful in screening compounds for effectiveness as agonist  
CC or antagonist of the polypeptides, or in altering the expression of the  
CC target polynucleotide and compounds that specifically bind to or modulate  
CC the activity of the polypeptide. The microarray is useful in monitoring  
CC or measuring protein-protein interactions, drug-target interactions, and  
CC gene expression profiles. This polynucleotide sequence represents the DNA  
CC encoding a human PMOD protein of the invention  
XX  
SQ Sequence 2572 BP; 780 A; 489 C; 509 G; 794 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 7,47e-132 Length: 2572  
Score: 1240.00 Matches: 229  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 8 Gaps: 0  
  
US-10-649-273-2\_COPY\_176\_414 (1-239) x ABT23207 (1-2572)  
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIle 20  
DB 669 CTGTTGGCATTAAGTTCAAGAGAGTTTCACATTTTCTGCTTGGAAAGCTTTGGACATA 728  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgGluSerLeuIleLysHisProGlu 40  
DB 729 GCACCAAGTGCATGCTTGCACAGGTGCAAGAACATTTCTTATTAATMAACATCCAGG 788  
QY 41 CysSerThrMetSerGlyGlyLysValAlaIleGluHisLeuAlaLysGlnGlyValAsnArgPhe 60  
DB 789 TGTCTCACCATGAGTGGGGAAGCCATAGAACATTTGGCCAAACAGAAATAGATT 848  
QY HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
DB 849 CATTTTGACATCAAAACCTCCCTTGATCATCTAATAAATTGATTTTCTTTACTGA 908  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100  
DB 909 CTTCAACAGCTTAATGATTAATTAATTAAGAAAAAGAAAGAGAGATTGAGAAAG 968  
QY 101 GlyGlnIleLeuSerSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetLys 120  
DB 969 GGGCAAAATCCGTCTTTCAGCAGACACTTCTGCACAGTACACACACCAATGCAATG 1028  
QY HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGluArgAspLeuLeuPro 140  
DB 1029 CATCTTGGAAAGAACATCGGCTATTCTGTTTGTAAACAGAGACTTGTACT 1088  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaAspAsnPhenylleArgArg 160  
DB 1089 CAAATATATGACAGTACTGTTGATCTGGTGTCGCCAAGTAACTTCTATATCCGAGCA 1148



QY 161 AlAlenGluileuThrasnaAlaThrGlnCysThrleuLeuCysProProProArgLeu 180  
DB 1149 GCTCTGGAATTTTAAACAAACGACACAGTGCATTTGTGTGTCTCTCCGACACTA 1208  
QY 181 CysThraspaSnglylleMetlleAlaThrasGlylleGluArgLeuAlaGlyLeu 200  
DB 1209 TGCACGTATATATGCAATATGATGCAATGCAATGATTAACACATACGCTGGCTTG 1268  
QY 201 GlylleleuHisAspilleGluGlylleArgTyrgLupProLysCysProLeuGlyValAsp 220  
DB 1269 GGCATTTTACATACATACAGGCAATCCGCTATACACCAAAATCTCTTGGAGTGAAC 1328  
QY 221 lleserlySGluValGlyGluAlaSerlleLysValProGlnLeuLysMetGluile 239  
DB 1329 ATATCAAAAGAACTTGAGAAAGCTTCCATAAAGTACCAACATTTAAATGAGATA 1385  
RESULT 6  
AAD46856  
ID AAD46856 standard; cDNA; 1820 BP.  
AC AAD46856;  
XX 27-JAN-2003 (first entry)  
XX Human glycoprotease 28472 cDNA.  
DE  
XX Human; adenosine deaminase; seven transmembrane domain receptor; Cancer;  
KW 77M; glycoprotease; immune disorder; IGA deficiency; allergy; arrhythmia;  
KW rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;  
KW hyperextension; ischemic heart disease; obesity; myocardial infarction;  
KW endothelial cell disorder; Grave's disease; psoriasis; brain disorder;  
KW Parkinson's disease; Alzheimer's disease; haematopoietic disorder;  
KW cerebral oedema; metabolic disorder; liver disorder; platelet disorder;  
KW chromosome mapping; tissue typing; gene therapy; neuroprotective;  
KW cytosolic; anorectic; cardiac; haemostatic; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH 146..1390  
FT CDS /\*tag= a  
FT /product= "Human 28472 protein"  
FT /note= "This region is specifically claimed as SEQ ID NO:  
FT 6 in claim 1 of the specification"  
XX  
XX WO200274960-A2.  
XX  
XX 26-SEP-2002.  
XX  
XX 08-NOV-2001; 2001WO-US051427.  
XX  
XX 08-NOV-2000; 2000US-0246768P.  
XX 08-NOV-2000; 2000US-0246772P.  
XX 15-NOV-2000; 2000US-0249185P.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Leiby KR, Kapeller-Libermann R, Glucksmann M;  
PI P-PSDB; AAE29234.  
DR  
XX MPI; 2002-759898/82.  
XX  
XX P-PSDB; AAE29234.  
XX  
XX New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,  
XX useful for diagnosing and treating cancer, immune, cardiovascular,  
XX PT hematopoietic, brain, pain, metabolic, liver or platelet disorders, and  
XX PT in pharmacogenomics.  
XX  
XX Claim 1, Fig 8, 178pp; English.  
XX  
XX The present invention relates to novel 38650, 28472, 5495, 65507, 81588  
XX CC or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-  
XX CC protease or seven transmembrane domain (77M) receptor family members.  
XX CC Sequences of the invention are useful in diagnosing and treating cancer

CC or aberrant cellular proliferation and/or differentiation (e.g. colon or  
CC lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid  
CC arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,  
CC hyperextension, atherosclerosis, arrhythmias, ischemic heart disease,  
CC myocardial infarction, thrombus) including endothelial cell disorders  
CC (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain  
CC disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),  
CC pain and metabolic disorders (e.g. obesity), liver disorders or platelet  
CC disorders. They are also useful in screening assays, predictive medicine  
CC (e.g. diagnostic assays, prognostic assays, monitoring clinical trials  
CC and pharmacogenetics) and prophylactic and therapeutic methods. The  
CC nucleic acids may also be used in chromosome mapping, tissue typing and  
CC forensic biology and as surrogate markers. Sequences of the invention are  
CC also used in gene therapy. The present sequence is human glycoprotease  
CC 28472 cDNA  
XX  
SQ Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;  
Alignment Scores:  
Pred. No.: 8.22e-128 Length: 1820  
Score: 1203.00 Matches: 232  
Percent Similarity: 98.33% Conservative: 3  
Best Local Similarity: 97.07% Mismatches: 4  
Query Match: 97.02% Indels: 0  
DB: Gaps: 0  
US-10-649-273-2\_copy\_176\_414 (1-239) x AAD46856 (1-1820)  
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 671 CTGTGGCATATGATGTTCAAGGAGGTTTCACATTTTCCTTCTTGGAAGCTTTGACATTA 730  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuLysHisProGlu 40  
DB 731 GCACCAAGGTGACATGCTTGACAGGTGCAAGAAACATCTTTTATTAACATCCAGAG 790  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60  
DB 791 TGCTCCACCATGATGATGGTGGGAAAGCCATAGACATTTGGCCAAACAAATATGATTT 850  
QY 61 HisPheAspIleLysProProLeuHisAlaLysAsnGlyAspPheSerPheThrGly 80  
DB 851 CATTTTGACATCAAAACCTCCCTGCATCATCTAAATTTGATTTTCTTTTACTGGA 910  
QY 81 LeuGlnHisValThrAspLysIlelleMetLysLysGlnGlyIleGluLys 100  
DB 911 CTTCAACAGTTACTGATTAATAATATGAAAAACAGAAACAGATTTTGAAG 970  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120  
DB 971 GGGCAAAATCTCTCTTCACGACAGACATTTGCTGCACAGTACACACATATGCAAT 1030  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
DB 1031 CATCTTGGAAGAAACATCGGGCTATTCTGTTTGTAAAGCAGAGACTTTGTAACT 1090  
QY 141 GluAsnAsnAlaValLeuValAlaSerGlyValAlaAsnPheTyrlleArgArg 160  
DB 1091 CAAATATATGCACTGATCTGTGCATCTGTGTGTGCGAAAGTATCTATATCCGCACA 1150  
QY 161 AlAlenGluileuThrasnaAlaThrGlnCysThrleuLeuCysProProProArgLeu 180  
DB 1151 GCTCTGGAATTTTAAACAAACGACACAGTGCATTTGTGTGTCTCTCCGACACTA 1210  
QY 181 CysThraspaSnglylleMetlleAlaThrasGlylleGluArgLeuAlaGlyLeu 200  
DB 1211 TGCACGTATATGCAATATGATGCAATGCAATGATTAACACATACGCTGGCTTG 1270  
QY 201 GlylleleuHisAspilleGluGlylleArgTyrgLupProLysCysProLeuGlyValAsp 220  
DB 1271 GGCATTTTACATACATACAGGCAATCCGCTATACACCAAAATCTCTTGGAGTGAAC 1330  
QY 221 lleserlySGluValGlyGluAlaSerlleLysValProGlnLeuLysMetGluile 239

DB 1331 ATATCAAAAGAGTTGGAGAGCTTCATTAAGTACCACTTAATAATGAGATA 1387  
|||||  
RESULT 7  
ACAG0887  
ID ACAG0887 standard; cDNA; 1820 BP.  
XX  
AC AG0887;  
XX  
DT 08-JUL-2003 (first entry)  
XX  
DE Human cDNA 28472 encoding a glycoprotease.  
XX  
KW Human; se; gene; cancer; aberrant cellular proliferation;  
KW differentiation; immune disorders; heart disorder; brain disorder;  
KW cardiovascular disorder; endothelial cell disorder; pain disorder;  
KW haematopoietic disorder; blood vessel disorder; metabolic disorder;  
KW liver disorder; platelet disorder; glycoprotease.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 146..1390  
FT /tag=a  
FT /product="Glycoprotease"  
FT /note="This CDS is specifically claimed in claim 1"  
XX  
US200309017-A1.  
XX  
PD 09-JAN-2003.  
XX  
PF 08-NOV-2001; 2001US-00012140.  
XX  
PR 08-NOV-2000; 2000US-0246768P.  
PR 08-NOV-2000; 2000US-0246772P.  
PR 15-NOV-2000; 2000US-0249185P.  
XX  
PA (LEIB/) LEIBY K R.  
PA (KAP/) KAPPELLER-LIBERMANN R.  
PA (GLUC/) GLUCKSMANN M A.  
XX  
PI Leiby KR, Kapeller-Libermann R, Glucksmann MA;  
XX  
DR WPI; 2003-428888/40.  
DR P-SDB; AB09559.  
XX  
PT New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid  
XX molecules, useful for diagnosing, treating cancer, pain, or immune,  
XX heart, endothelial cell, haematopoietic, blood vessel, brain, metabolic  
XX and liver disorders.  
XX  
PS Claim 2; Fig 8; 90pp; English.  
XX  
XX The invention relates to an isolated 38650 (encoding adenosine  
XX deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7  
XX transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or  
XX a sequence which is at least 60% identical to the six nucleic acids or  
XX their open reading frames, fragments of at least 15 nucleotides,  
XX naturally occurring variants, or a DNA insert of the plasmid deposited  
XX with the American Type Culture Collection as Accession No. not defined in  
XX the specification, which encodes the amino acid sequence). Also included  
XX are a host cell containing the nucleic acid (used to produce the  
XX proteins), the encoded proteins, an antibody that selectively binds to  
XX the polypeptide, and identifying a compound that binds to/modulates the  
XX activity of the polypeptide. The nucleic acid molecules, polypeptides and  
XX methods are useful for diagnosing, treating cancer, aberrant cellular  
XX proliferation and/or differentiation, immune disorders, heart disorders,  
XX cardiovascular disorders including endothelial cell disorders,  
XX haematopoietic disorders, blood vessel disorders, brain disorders, pain  
XX and metabolic disorders, liver disorders and platelet disorders (many  
XX examples of these disorders are given in the specification). The present  
XX sequence is the Human cDNA 28472 encoding a glycoprotease

SO Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;  
Alignment Scores:  
Pred. No.: 8, 22e-128 Length: 1820  
Score: 1203.00 Matches: 232  
Percent Similarity: 98.33% Conservative: 3  
Best Local Similarity: 97.07% Mismatches: 4  
Query Match: 97.02% Indels: 0  
DB: 8 Gaps: 0  
US-10-649-273-2\_COPY\_176\_414 (1-239) x ACAG0887 (1-1820)  
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 671 CTTGGTGGCATTTAGTTCAAGAGATTTCAGATTTCTCTTGGAAAGCTTTGGACATA 730  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuLysHisProGlu 40  
DB 731 GCACCAAGTGACATGCTTGACAAAGGTGGCAAGAGACTTTCTTAATAAACATCCAGAG 790  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
DB 791 TGTCCACCATAGAGTGGTGGAAAGCCATAGAACATTTGGCCAAACAGGAATAAGATT 850  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
DB 851 CATTTTGACATCAAACTCTCTTGATCATGTGTAATAATTTGATTTCTTTACTGGG 910  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100  
DB 911 CTTCAACAGCTTACGATTAATAATATATGAAACAGGAAACAGAGAGATTTGGAAG 970  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120  
DB 971 GGGCAATTCCTGTCTTCAGCAGAGCATTTGCTCCACGATACAGACCAATGCGATGT 1030  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
DB 1031 CATCTTGGAAGAAAGAACATATGGGGTATTCCTTTTGTAGAGAGAGACTTTGTAACCT 1090  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160  
DB 1091 CAAATTAATGCAGTCTGCTGATCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1150  
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGlu 180  
DB 1151 GCTCTGGAATTTTAACAAACCAACACAGTGCATTTGTGTCTCTCCCAACATCA 1210  
QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
DB 1211 TGCACCTGATTAATGCAATTTGATTTGATGGAATGATTTGGAAGACTACGCTGCTTG 1270  
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgIleArgIleProLysCysProLeuGlyValAsp 220  
DB 1271 GGCAATTTACATGATGATGAAAGCATCCGCTATGAAACAAATATGCTCTTGGAGTAGAG 1330  
QY 221 HisLeuGlyGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239  
DB 1331 ATATCAAAAGAGTTGGAGAGCTTCATTAAGTACCACTTAATAATGAGATA 1387  
RESULT 8  
ABSS7020  
ID ABSS7020 standard; cDNA; 1821 BP.  
XX  
AC ABSS7020;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE cDNA encoding novel human glycoprotease 28472.  
XX  
KW Cancer; aberrant cell proliferation; aberrant cell differentiation;  
KW breast cancer; ovarian cancer; prostate cancer; colon cancer;  
KW lung cancer; immune disorder; heart disorder; cardiovascular disorder;



PR 24-JAN-2002; 2002US-0350435P.  
XX (HELI-) HELIX RES INST.  
PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
PI Isegai T, Sugiyama T, Otsuki T, Makamatsu A, Sato H, Ishii S;  
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;  
XX WPI; 2003-395539/38.  
DR P-PSDB; ADA54471.  
XX  
PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
PT and/or membrane proteins, useful for developing medicines for diseases in  
PT which the gene is involved, or as target molecules for gene therapy.  
XX  
PS Claim 1; SEQ ID NO 400; 205bp; English.  
XX  
CC The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
SQ Sequence 2208 BP; 660 A; 454 C; 451 G; 643 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 3,88e-111 Length: 2208  
Score: 1059.00 Matches: 211  
Percent Similarity: 88.28% Conservative: 0  
Best Local Similarity: 88.28% Mismatches: 4  
Query Match: 85.40% Indels: 24  
DB: 10 Gaps: 1  
  
US-10-649-273-2\_COPY\_176\_414 (1-239) x ADA52832 (1-2208)  
  
QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
Db CTGTTGGCATTAGTTCAAGAGAGTTTCACATTTTCCTTGGAAAGCTTTGGACATA 928  
  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
Db GCACACAGGTGACATCTTGCACAGGTGCAAGACATCTTATATAAATCCAGAG 988  
  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaValGInGlyLysAspPhe 60  
Db TGCTCCACCATGAGGGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATGATTT 1048  
  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
Db CATTTTGACATCAAACTCCCTTGATCATGCTAATAAATTGATTTCTTTACTGCA 1108  
  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100  
Db CTTCAACACGTTACTGATTAATAATATATGAAAAAGAAAAAGAAAGATTTGAGAG 1168  
  
QY 101 GlyValIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetLacys 120  
Db GGGCAAAATCTGTCTTGACAGACACATTCCTGCACAGTACACACACAAATGCATCT 1228  
  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
Db CATCTTGTGAAGAAGACATCGGCTATTCTGTTTGTAGACAGAGACTGTGTTACTT 1288  
  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaAlaSerAsnPheTyrIleArgArg 160  
Db CAAATATATGACAGTACTGTTGCATCTGTGTGTGCGAAGTAACTTCTGTATCGGACA 1348  
  
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180  
Db GCTCTGGAAATTTTAAACAAACGCAACAGTGCATTTGTGTGTGTCCTCTCCAGACTA 1408  
  
QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200

Db 1409 TGCATCTGATTAATGCACTTAATGATTGCA----- 1435  
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220  
Db 1436 -----TGAATCTCTCTTGGAGTAC 1456  
QY 221 HisSerLysGluValAlaGlyAlaSerIleLysValProGlnLeuLysMetGluIle 239  
Db 1457 ATATCAAAAGAAAGTGCAGAAAGCTTCATTAAGTACCAATTAATAATGAGATTA 1513  
  
RESULT 10  
ADQ24627  
ID ADQ24627 standard; DNA; 2890 BP.  
XX  
AC ADQ24627;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7447.  
XX  
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
XX ds.  
XX  
OS Homo sapiens.  
XX  
PN WO2004048938-A2.  
XX  
PD 10-JUN-2004.  
XX  
PF 26-NOV-2003; 2003WO-US038193.  
XX  
PR 26-NOV-2002; 2002US-0429739P.  
XX  
PA (PROT-) PROTEIN DESIGN LABS INC.  
XX  
PI Aziz N, Ginsburg WM, Zlotnik A;  
XX  
DR WPI; 2004-441208/41.  
XX  
PT Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX  
PS Example 2; SEQ ID NO 7447; 210bp; English.  
XX  
CC The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX  
SQ Sequence 2890 BP; 869 A; 609 C; 611 G; 789 T; 0 U; 12 Other;  
  
Alignment Scores:  
Pred. No.: 5.67e-111 Length: 2890  
Score: 1059.00 Matches: 211  
Percent Similarity: 88.28% Conservative: 0  
Best Local Similarity: 88.28% Mismatches: 4  
Query Match: 85.40% Indels: 24  
DB: 12 Gaps: 1  
  
US-10-649-273-2\_COPY\_176\_414 (1-239) x ADQ24627 (1-2890)

QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 1526 CTGTTGGCATATGATTCAGAGAGTTTCAGATTTTCGCTTCCTGGAAGCTTTGGACATA 1585  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgAspGluSerLeuIleLysHisProGlu 40  
DB 1586 GCACACAGGTGACATGCTTGAACAGGTGCAGAGAACATCTCTTATATATAACATCCAGG 1645  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGluGlyAsnArgPhe 60  
DB 1646 TGTCTCCACATGATGATGATGGGAAAGCCATAGAACCTTTGGCCAAACAGAAATATGATTT 1705  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80  
DB 1706 CATTTTGACATCAACCTCCCTTGACATCATGCTAAATATGATTTTCCTTTTCTGGA 1765  
QY 81 LeuGlnHisValIleThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100  
DB 1766 CTTCACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1825  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120  
DB 1826 GGGGCAATCTGCTCTTCAGACAGACATGCTGCAAGTACAGACACAAATGGCATGT 1885  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
DB 1886 CATTTTGGAAAGAACACATCGGGCTATTCCTGTTTGGTAAACAGAGACTGTTTAACT 1945  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaLysSerAsnPheTyrIleArgArg 160  
DB 1946 CAAATATATGACGTCTGCTGTCATCTGTCATCTGTCGCAATGATCTGATCCGACGA 2005  
QY 161 AlaLeuGluIleLeuThrAspAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180  
DB 2006 GCTCTGAAATTTTAAACAAACGACACACATGCTGTTGTTGTCCTCTCCACAGACTA 2065  
QY 181 CysThrAspAsnGlyIleMetIleAlaTyrAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
DB 2066 TGCACTGATATGCAATATGATTTGCA----- 2092  
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220  
DB 2093 -----TGATGTCCTCTTGAGAGAGAC 2113  
QY 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239  
DB 2114 ATATCAAAAGAAAGTTGAGAGAGCTTCATTAAGTACCAAAATTAAGTGAAGATA 2170  
RESULT 11  
ADE31345/c  
ID ADE31345 standard; DNA, 3358 BP.  
XX  
AC ADE31345;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human diagnostic and therapeutic polynucleotide (dithp), SEQ ID NO 100.  
XX  
KW diagnostic and therapeutic polynucleotide; dithp; antiarteriosclerotic;  
KW antiinflammatory; cerebroprotective; antidiabetic; antidiabetic;  
KW immunosuppressive; neuroprotective; nootropic; neurologic; cranulizer;  
KW osteopathic; antiarthritic; antirheumatic; cytostatic; hepatotropic;  
KW vitamin; haemostatic; anti-HIV; antithyroid; thyromimetic;  
KW dermatological; antibacterial; fungicide; antiparasitic; anticonvulsant;  
KW chromolytic; anticoagulant; anorectic; vasorelaxant; antidiabetic;  
KW gene therapy; protein replacement therapy; human; gene; de.  
XX  
OS Homo sapiens.  
XX  
PN WO2003062376-A2.  
XX  
PD 31-JUL-2003.  
XX

PF 13-JAN-2003; 2003WO-US001096.  
XX  
PR 16-JAN-2002; 2002US-0349384P.  
PR 17-JAN-2002; 2002US-0349413P.  
PR 17-JAN-2002; 2002US-0349946P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
PI Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JH,  
PI Yu JY, Tuson O, Yap PE, Amshay SR, Dam TC, Liu TF, Gerstein EH,  
PI Peralta CH, Lewis SA, Chen A, Marwaha R, Ian RV, Urashka ME,  
PI Kristnam SR, Kolluru V, Panesar IS;  
DR WPI; 2003-636732/60.  
DR P-PSDB; ADE31156.  
XX  
PT New human diagnostic and therapeutic polynucleotides and polypeptides,  
PT useful for diagnosing, treating or preventing e.g. leukemia, brain  
PT cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke  
PT or Alzheimer's.  
XX  
PS Claim 1, SEQ ID NO 100; 634pp; English.  
XX  
CC The invention relates to a novel isolated human diagnostic and  
CC therapeutic polynucleotide (designated dithp). The novel dithp  
CC polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798  
CC base pairs fully defined in the specification; a polynucleotide  
CC comprising a naturally occurring polynucleotide sequence at least 90%  
CC identical to the dithp polynucleotide; a polynucleotide complementary to  
CC the dithp polynucleotide or its polynucleotide which is at least 90%  
CC identical; or an RNA equivalent of any of the polynucleotides mentioned  
CC above. The dithp polynucleotides have the following activities:  
CC antiarteriosclerotic, antiinflammatory, cerebroprotective, antidiabetic,  
CC antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,  
CC cranulizer, osteopathic, antiarthritic, antirheumatic, cytostatic,  
CC hepatotropic, vitamin, haemostatic, anti-HIV, antithyroid, thyromimetic,  
CC dermatological, antibacterial, fungicide, antiparasitic, anticonvulsant,  
CC thrombolytic, anticoagulant, anorectic, vasorelaxant, and antidiabetic. The  
CC novel dithp polynucleotides polypeptide can be used in gene therapy and  
CC protein replacement therapy. The dithp polynucleotides or dithp  
CC polypeptides are useful for diagnosing, preventing or treating diseases  
CC associated with the expression of human molecules. In particular, these  
CC diseases include cancers (e.g. adenocarcinoma, leukemia, melanoma, brain  
CC cancer, breast cancer, cervix cancer, bone cancer, liver cancer, lung  
CC cancer) or other cell proliferative disorders (e.g. arteriosclerosis,  
CC atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary  
CC chromocytopenia), autoimmune/inflammatory disorders (e.g. AIDS,  
CC Addison's disease, thyroiditis, Crohn's disease, Graves' disease,  
CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid  
CC arthritis, scleroderma, systemic lupus erythematosus), infections (e.g.  
CC viral, bacterial, fungal or parasitic infection), developmental disorders  
CC (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.  
CC chromoblasts, hypopituitarism, hypogonadism, gigantism, goiter) metabolic  
CC disorders (e.g. hypercholesterolaemia, hypoglycaemia, diabetes,  
CC hyperlipidaemia, obesity), neurological disorders (e.g. ischaemic  
CC cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,  
CC Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease,  
CC anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers),  
CC transport disorders (e.g. akinesia or multidrug resistance), or  
CC connective tissue disorders (e.g. Paget's disease or rickets). This  
CC polynucleotide sequence represents one of the human dithp DNA sequences  
CC of the invention.  
XX  
SQ Sequence 3358 BP; 1105 A; 577 C; 601 G; 1075 T; 0 U; 0 Other;  
XX  
Alignment Scores:  
Pred. No.: 1,4e-103 Length: 3358  
Score: 995.50 Matches: 209  
Percent Similarity: 87.82% Conservative: 0  
Best Local Similarity: 87.82% Mismatches: 2  
Query Match: 80.28% Indels: 29  
DB: 10 Gaps: 1

US-10-649-273-2\_COPY\_176\_414 (1-239) x ADB31345 (1-3358)

QY 29 ValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGly-GlyLy 48

Db 2104 GTGGCAAGAAAGACTTTCTTATATAAACATCCAGAGTGCTCCACCATGATGCTGGGAA 2045

QY 48 GAlaIleGluHisLeuAlaLysGlnGlyValAsnArgPheHisPheAspIleLysProProle 68

Db 2044 AGCCATAGAAACATTTGGCCAAACAGAAATAGATTTCATTTTGACATCAACCTCCCTT 1985

QY 68 uHisHisAlaLysAsnGlyAspPheSerPheThrGlyLeuGlnHisValThrAspLysIle 88

Db 1984 GCATCATGCTTAAATATGTAATTTCTTTTCTTGAAGCTTCAACAGCTTACTGATTAAT 1925

QY 88 eIleMetLysLysGluGlyGluGly-----GlyIleGluLysGlyGlnI 96

Db 1924 AATATATGAAAAAGAAAAAGAAAGATATTTCTAATTAAGTAAAGTAAAGATAA 1865

QY 97 -----GlyIleGluLysGlyGlnI 103

Db 1864 TATTCCTGATTTGCTTAAATAATAGCTCTCATTTCTGACAGGTATTGGAAGGGGAA 1805

QY 103 IleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisPheMetAlaCysHisLysLys 123

Db 1804 TCTGTCTTACGACGACAGACATTCCTGACACAGTACACACATGCGATGCTCATTTG 1745

QY 123 allYsAQTThHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsn 143

Db 1744 TGAAAAAACAACATCGGCTATTTCTTTTGTAAACAGAGACTTTTACTCTCAAAATA 1685

QY 143 snAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuG 163

Db 1684 ATGCACTACTGCTTGCATCTGT-GGTGTGCAAGTAACTTCTATATCCGACAGCTCTGG 1626

QY 163 IuIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuGlyThr 183

Db 1625 AAATTTTAAAC-AAAGCAACACAGTGCATTTGTGTGCTCTCCAGCATATCACCTG 1567

QY 183 sPaAngIyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIle 203

Db 1566 ATAATGGCATTAATGATTCATGAGAAATGATTGAAGACTACGCTGCTGGGCAATTT 1507

QY 203 euHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSer 223

Db 1506 TACATGACATGAGAGGATCGGCTATGAAACCAAAATGCTCTTGAGATGACATATCA 1447

QY 223 ysgIuValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239

Db 1446 AAGAAGTTGGAGAGCTTCCATAAAGTACCAATTAATAATGAGATA 1397

RESULT 12

AB075508 standard; DNA; 1572 BP.

AC AB075508;

XX 07-NOV-2002 (first entry)

XX Murine sialoglycoprotease-like gene sequence SEQ ID NO:7.

XX Murine; mouse; protease; calcium activated neutral protease type 5;

KM CAPN5; trypsin 4; sialoglycoprotease; enzyme; genetic disease;

KM neurological; neuropsychological; psychotic illness; transgenic animal;

XX gene; ds.

OS Mus musculus.

XX PN M0200245491-A2.

XX 13-JUN-2002.

XX 05-DEC-2001; 2001WO-US046405.

PR 06-DEC-2000; 2000US-0251803P.

PR 06-DEC-2000; 2000US-0251820P.

PR 13-DEC-2000; 2000US-0255971P.

PA (DELT-) DELTAGEN INC.

XX Allen KD, Leviten MW;

XX WPI; 2002-657389/70.

PT Novel transgenic animal, comprising a disruption in protease target gene,

PT with a disruption in a protease target gene.

XX Example 3; Fig 7; 62pp; English.

CC The present invention describes a non-human transgenic animal (I)

CC comprising a disruption in a protease target gene (PG) selected from

CC calcium activated neutral protease type 5 (CAPN5) gene, trypsin 4 gene

CC and sialoglycoprotease-like gene. Also described is a targeting construct

CC (II), comprising a first polynucleotide sequence homologous to at least a

CC first portion of PG, a second polynucleotide sequence homologous to at

CC least a second portion of PG and a selectable marker. (II) is useful for

CC producing a transgenic mouse comprising a disruption in a protease target

CC gene, by introducing (II) into a cell, introducing the cell into a

CC blastocyst, implanting the resulting blastocyst into a pseudopregnant

CC mouse, where the pseudopregnant mouse gives birth to a chimeric mouse,

CC and breeding the chimeric mouse to produce the transgenic mouse. (I) is

CC useful for identifying an agent that modulates the expression or function

CC of a protease target gene, by administering an agent to (I) and

CC determining whether the expression or function of the disrupted protease

CC target gene in (I) is modulated. (I) is also useful for testing the

CC efficacy of proposed genetic and pharmacological therapies for human

CC genetic diseases, such as neurological, neuropsychological or psychotic

CC illnesses. The present sequence represents murine sialoglycoprotease-like

CC gene sequence, which is used in an example from the present invention

XX

SQ Sequence 1572 BP; 459 A; 337 C; 340 G; 429 T; 0 U; 7 Other;

Alignment Scores:

Pred. No.: 1.28e-89 Length: 1572

Score: 870.00 Matches: 184

Percent Similarity: 82.77% Conservative: 13

Best Local Similarity: 77.31% Mismatches: 37

Query Match: 70.16% Indels: 5

DB: Gaps: 2

US-10-649-273-2\_COPY\_176\_414 (1-239) x AB075508 (1-1572)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20

Db 528 CTGTGGKRTTATGCTCAAGAGTGTTCCTGATTCCTCTTGGAAGTCTTTGACATA 587

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40

Db 588 GCACACAGGCGCATGCTTGAACAAGTGGCAAGAGCTTTCTTATTAACAACATCCAGA 647

QY 41 CysSerThrMetSerGlyGlyLysValAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60

Db 648 TGTTCACAAATGAGTGTGGAAGACTATAGAACAGTTGCGCCAAACGGAATAATGATTC 707

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80

Db 708 CATTTTACTATCAATCCATTCATGCAAGATGCTTAAGAAATTTGCAATTTTCTTCACGGGA 767

QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlyLysGlnGlyIleGluLys 100

Db 768 CTTCAACATTAATTAAGTAACTAATTAACACACAGAAAAAGAAAGCATTTGAGAG 827

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120

Db 828 GGGCAATCTCTGTATCACTGACGTCAGACATTTGCTGCGGTACAGATCAACAGCGTGC 887



QY 121 HisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuLeuPro 140  
CC |||||  
CC CACCTTCGAAAGAAAGACATCGCGCTATTCGTTTTCAGACAGAAAAATTTGCTCT 947  
Db 888  
QY 141 GlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160  
CC |||||  
CC 948 CCAAGTAAACCAAGTAAAGTGTGTCTGAGAGTGTGCAAGTAACTGTACATCCGAAA 1007  
Db 161 AlLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180  
QY |||||  
Db 1008 GCATTGGAATTTGTCGCAATATGCAACGATCGACGCTTTGTGTCTCACT-TCAGACATG 1066  
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
Db 1067 TGCACCTGACATGCGCATATGATTCAGATGAAATGGAATTCAGTCCCTGCTT 1126  
QY 201 GlyIleLeu-HisAsp-IleGluGlyIleArgGlyIleArgProGlyCysProLeuGlyVal- 219  
Db 1127 GGCCTTTTACCATGATGATGATGAGACATTCGGTTATTAACCAAAATGTCCTCTTGAGTGA 1186  
QY 220 --AspIleSerIysGluValGlyGluAlaSerIleIysValProGlnLeu 235  
Db 1187 GGCATTATCCCGAAGAGTTGGCAGA--AGCTTGCCCATTTAAAAAGTTA 1233  
RESULT 13  
ID AAS84622 standard; cDNA; 2734 BP.  
AC AAS84622;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #20426.  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
OS Homo sapiens.  
XX  
PN W0200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PE 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Dirmacac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG20435.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 1; SEQ ID NO 20426; 103pp; English.

CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2734 BP; 773 A; 545 C; 639 G; 763 T; 0 U; 14 Other;  
Alignment Scores:  
Pred. No.: 4.02e-58 Length: 2734  
Score: 599.00 Matches: 131  
Percent Similarity: 71.57% Conservative: 10  
Best Local Similarity: 66.50% Mismatches: 22  
Query Match: 48.31% Indels: 34  
DB: 5 Gaps: 6  
US-10-649-273-2\_COPY\_176\_414 (1-239) x AAS84622 (1-2734)  
QY 43 ThrMetSerGlyGlyValAlaIleGluHisLeuAlaIysGlnGlyAsnArgPheHisPhe 62  
Db 206 ACCCTGCTAGGGGGGCGCCGCT-----AGATTCCAT----- 235  
QY 63 AspIleIysProProLeuHisHisAlaIys-----AsnCys----- 74  
Db 236 -----CCTATTCTCCCGATGAAAGTATCAGGCTGTGTAGCAAGCGTGGACGCA 283  
QY 75 -----AspPheSerPheThrGlyLeuGlnHisValThrAsp 86  
Db 284 TCTGCTTCGGGGAGGCCGCCAGGAGCTTTTACTCATCGCGAAGCAAAATTGGAGATAG 343  
QY 87 IysIleIleMetIysIysGlyIysGluIysGluIys-----GlyIleGluIysGlyIleLeu 104  
Db 344 CGCTTACATGACGAGACGAGACGAGACGAGGGGGGCTATTGAGAAAGGGGCAAACTCTG 403  
QY 105 SerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValIys 124  
Db 404 TCTTACGACGACGACATTCCTGCTCCACATGACGACACATGACATGCTATCTTGAGAA 463  
QY 125 ArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuProGlnAsnAla 144  
Db 464 AGAACATCGGGCTATTCTGTTTGTATAGCAGAGACTTGTACTCTCAAAATATGCA 523  
QY 145 ValLeuValAlaSerGlyGlyValAlaIleSerAsnPheTyrIleArgArgAlaLeuGluIle 164  
Db 524 GTACTGTTGCACTGCTGTGTGTGCGCACTTCTATATCCGACAGCTCGAAAT 583  
QY 165 LeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsn 184  
Db 584 TTAACAAACGCAACACAGTCACTTGTGTGTCTCTCCACGACTTATGCACTGATTAAT 643  
QY 185 GlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHis 204  
Db 644 GCATTATTAATAATGCAATGATGTATTTGAAAGACTACGTCGTGGCTTGGCATTTTACAT 703  
QY 205 AspIleGluGlyIleArgGlyIleArgProIys-----CysProLeuGlyVal 219  
Db 704 GACATTAAGAGGCAATCCGCTATGAAACCAAGTATGTGGCTCTTCAGAGGCTG 754  
RESULT 14  
ID ADL86725  
ADL86725 standard; DNA; 371 BP.  
AC ADL86725;  
XX  
DT 20-MAY-2004 (first entry)  
XX  
DE DNA up-regulated in murine multipotent progenitor cells Segid 3118.  
XX



XX	gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
KW	HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
KM	common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds
OS	Mus sp.
XX	
PN	WO2003093445-A2.
XX	
PD	13-NOV-2003.
XX	
PF	05-MAY-2003; 2003WO-US014114.
XX	
PR	03-MAY-2002; 2002US-0377383P.
XX	
PA	(STOW-) STOWERS INST MEDICAL RES.
XX	
PT	Lt L;
XX	
DR	WPI; 2004-022656/02.
XX	
PT	Classifying an unknown multi-lineage affiliated gene comprises isolating
XX	expressed nucleic acid sequences from the discrete cell sub-populations.
XX	
PS	Claim 8; SEQ ID NO 3118; 123p; English.
XX	
CC	This invention relates to a novel method for predicting gene potential by
CC	associating nucleic acid sequences of unknown function with particular
CC	sub-population profiles. Specifically, it refers to classifying an
CC	unknown multi-lineage affiliated gene by collecting hybridisation data to
CC	develop a gene expression map, in order to determine the discrete sub-
CC	population where it is expressed. The present invention describes methods
CC	for predicting the lineage commitment of genes associated with the self-
CC	renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
CC	self renewing multipotent progenitors (MPPs), common lymphoid progenitors
CC	(CLPs) and common myeloid progenitors (CMPs), which are collectively
CC	referred to as bone marrow stem cells populations. As such, these methods
CC	can be used to identify associated multi-lineage affiliated genes and
CC	hence the underlying molecular mechanisms in physiological haematopoietic
CC	development. This polynucleotide sequence is DNA associated with a murine
CC	MPP sub population of cells of the invention.
XX	
SQ	Sequence 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other;
XX	
Alignment Scores:	
Pred. NO.:	2.75e-44
Score:	468.00
Percent Similarity:	78.86%
Best Local Similarity:	74.80%
Query Match:	37.74%
DB:	12
	Gaps: 0
US-10-649-273-2_COPY_176_414 (1-239) x ADL6725 (1-371)	
QY	109 AsplleAlaAlaThVaGlnhlsrMecAlaCyHlsleuValysArGThrhlsArG 128
DB	2 GACATGTCTGCTGCGGTACAGATGCACAGCGCTGCCACTTCGCAAAAGAACACATCCG 61
QY	129 AlAlleleuPheCyelsglnArGaPheleuPProGlnAsnAsnAlaValleuValAla 148
DB	62 GCTATTCTGTTTTCGACGACGAGAAATTTGCTCTCTCCAGCTAACCCAGATTAGTGTGA 121
QY	149 SerGjglYValAlaSerAsnPheryllleArGArGAlAleuGluTleuThrrAsnAla 168
DB	122 TCTGGAGGTGTTGCAAGTAACTTGTACATCCGAAAGCAATTGTCCGAAATGCA 181
QY	169 ThrGlnCystrleuLeuCyseProProArGArleuCyseThrAspAsnGlylleMetlle 188
DB	182 ACGCAGTGCACGTTGTGTGTCACCTCCAGACTGTGACATGCAATGACATCATGATT 241
QY	199 AlArTTrAnGjYrllleGlnArGleuArGAlaGlyleuGlyYlleuWlsAspIleGly 208
DB	242 GCATGGAAATGGAATTAAGATTACGTGACGCTTGGACCTTTTACATGAAATTAAGAAGC 301

```

Oy      209 11eatgTgTClupProlyeCySproueUGlyValAspIleSerlysgIuValGlyGuaLa 228
Db      302 ATCCNNATATGNAACCAAAATNTCNTMTTGAGTACATMTCCAGAGAA GTTGCAAGCT 361
Oy      229 SerIetIve 231
Db      362 GCCATPAAA 370

RESULT 15
ID      ADL86726
AC      ADL86726 standard; DNA; 371 BP.
XX
XX      ADL86726;
DT      20-MAY-2004 (first entry)
XX
XX      DNA up-regulated in murine multipotent progenitor cells SegID 3119.
DE
XX      gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
KW      HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
XX      common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; de-
XX
XX      Mus sp.
OS
XX      WO2003093445-A2.
XX      13-NOV-2003.
XX      05-MAY-2003; 2003WO-US014114.
XX      03-MAY-2002; 2002US-0377383P.
XX      (STOW-) STOWERS INST MEDICAL RES.
XX
XX      Li L;
XX
XX      WPI; 2004-022656/02.
DR
XX
XX      Classifying an unknown multi-lineage affiliated gene comprises isolating
PT      expressed nucleic acid sequences from the discrete cell sub-populations.
XX
XX      Claim 8; SEQ ID NO 3119; 123bp; English.
XX
XX      This invention relates to a novel method for predicting gene potential by
XX      associating nucleic acid sequences of unknown function with particular
XX      sub-population profiles. Specifically, it refers to classifying an
XX      unknown multi-lineage affiliated gene by collecting hybridisation data to
XX      develop a gene expression map, in order to determine the discrete sub-
XX      population where it is expressed. The present invention describes methods
XX      for predicting the lineage commitment of genes associated with the self-
XX      renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
XX      self renewing multipotent progenitors (MPPs), common lymphoid progenitors
XX      (CLPs) and common myeloid progenitors (CMPs), which are collectively
XX      referred to as bone marrow stem cells populations. As such, these methods
XX      can be used to identify associated multi-lineage affiliated genes and
XX      hence the underlying molecular mechanisms in physiological haematopoietic
XX      development. This polynucleotide sequence is DNA associated with a murine
XX      MPP sub population of cells of the invention.
XX
XX      SQ      Sequence 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other;

Alignment Scores:
Pred. No.:      2.75e-44      Length:      371
Score:      468.00      Matches:      92
Best Similarity:      78.86%      Conservative:      5
Best Local Similarity:      74.80%      Mismatches:      26
Query Match:      37.74%      Indels:      0
DB:      Gaps:      0

US-10-649-273-2_COPY_176_414 (1-239) x ADL86726 (1-371)
Oy      109 AsptleaaIaathvaIGlnhIeHrMetalaCyShIeIeuVallyaaRgThrniSarg 128

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: November 10, 2005, 16:53:13 ; Search time 3387.57 Seconds  
(without alignment)  
2685.516 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 L1ALVQGVSDPLLGKSLDI.....DISKEVGASIKVQLKMEI 239

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPFO\_spool\_p/US10649273/runat\_02112005\_091338\_15562/app\_query.fasta\_1.1429  
-DB=EST -QFMT=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human4.0.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273 @CGN 1 1 8076 @runat 02112005 091338 15562 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOC  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1118	90.2	870	5	BQ423651 AGENCOURT
2	1097	88.5	640	3	BQ636028 hnd03d11.y
3	1079	87.0	2284	3	AK045669 Mus muscu
4	1071	86.4	1622	3	AK011265 Mus muscu
5	930	75.0	852	5	BK391919 BX391919
6	926	74.7	1605	5	BK391919 BX391919
7	908	73.2	701	2	BCU30671 Mus muscu
8	876	70.6	658	7	BE740611 601595739
9	838	67.6	637	7	CF362328 829596 MA
					CK941819 4065407 B

10	799	64.4	922	5	BQ961028	BQ961028 AGENCOURT
11	789	63.6	545	1	AV602901	AV602901
12	784	63.2	490	6	CB852881	CB852881 UI-CF-FNO
13	783	63.1	822	7	CF257246	CF257246 pha008_90
14	753	60.7	866	5	BUI27463	BUI27463 603114407
15	749	60.4	736	7	CK365185	CK365185 AGENCOURT
16	746	60.2	723	3	BUI261251	BUI261251 603502215
17	731.5	59.0	792	5	BUI403563	BUI403563 604138456
18	717	57.8	634	2	AM601179	AM601179 RCL-BT025
19	708	57.1	736	6	CA057753	CA057753 esalrxb54
20	707.5	57.1	701	5	BUI621780	BUI621780 UI-H-FL1
21	698	56.3	696	5	BUI302605	BUI302605 603739448
22	685	55.2	537	6	CB272391	CB272391 ma157912
23	683	55.1	548	7	BK756548	BK756548 BX756548
24	682	55.0	548	7	CO880741	CO880741 Bowgen_09
25	671	54.1	682	2	BB043703	BB043703 CB43703
26	671	54.1	730	7	CN833245	CN833245 Oa.sp1bn
27	668	53.9	706	5	BUI202465	BUI202465 603949052
28	666	53.7	869	5	BK754527	BK754527 BX754527
29	647	52.2	919	7	CF407294	CF407294 CH3#047_E
30	646	52.1	484	1	AJ670918	AJ670918 AJ670918
31	646	52.1	878	5	BK776940	BK776940 BX776940
32	642	51.8	1173	6	CD508917	CD508917 CDA93-E05
33	638	51.5	1082	5	BK359023	BK359023 BX359023
34	625	50.4	506	2	BF415802	BF415802 UI-R-CA1-
35	614	49.5	861	5	BUI261505	BUI261505 603779906
36	611	49.3	1171	5	BUI261605	BUI261605 603501763
37	604	48.7	424	1	AA273921	AA273921 vb99c03_r
38	600	48.4	863	5	BUI376295	BUI376295 603808890
39	594	47.9	389	4	BM744822	BM744822 K-EST0018
40	591	47.7	812	5	BUI246489	BUI246489 603784202
41	589	47.5	357	5	BQ672554	BQ672554 AGENCOURT
42	581	46.9	704	5	BUI327284	BUI327284 603491570
43	576	46.5	826	7	CR444994	CR444994 CR444994
44	573	46.2	385	1	AJ647827	AJ647827 AJ647827
45	572.5	46.2	613	1	AA920105	AA920105 v225H05_r

#### ALIGNMENTS

RESULT 1  
BQ423651  
LOCUS BQ423651 870 bp mRNA linear EST 23-MAY-2002  
DEFINITION AGENCOURT 7790948 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6065828  
5', mRNA Sequence.  
ACCESSION BQ423651 GI:21118966  
VERSION BQ423651.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: ATCC/DC/DP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: L1AM13342 row: 1 column: 21  
High quality sequence stop: 710.

#### FEATURES

source  
1..870  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6065828"

```

/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 72"
/notes="Organ: Vector: PCMV-SPORE6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

```

## ORIGIN

## Alignment Scores:

```

Pred. No.: 3e-119      Length: 870
Score: 1118.00      Matches: 220
Percent Similarity: 98.22%      Conservative: 1
Best Local Similarity: 97.78%      Mismatches: 2
Query Match: 90.16%      Indels: 2
DB: 5      Gaps: 0

```

US-10-649-273-2\_COPY\_176\_414 (1-239) x BQ423651 (1-870)

```

Db      1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLeuSerLeuAspIle 20
      196 CTGTTGGCATTTGTTCAAGAGATTTCAGATTTCCTCTTGGAAGCTTTGGACATA 255
Qy      21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
      256 GCACAGGTGACATCTTGACAAAGGTGCAGAGAGACTTTCTTAATAAATCAATCCAGAG 315
Qy      41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaGlnGlyAsnArgPhe 60
      316 TGCTCCACCATGAGTGGTGGGAAAGCCATGAAACATTGGCCAAACAAAGAAATATATTT 375
Db      61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
      376 CATTTTGACATTAACCTCCCTTGATCTACTGCTAATAAATTTGATTTTCTTTACTGCA 435
Qy      81 LeuGlnHisValThrAspLysIleIleMetLysGlnLysGlnGlyIleGlnLys 100
      436 CTTCAACAGCTTACGATTAATAATTAATGAAAAAGAAAAAGAGAGATTTGAGAAAG 495
Db      101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
      496 GGGCAAAATCTGTCTTCAGCAGCAGACATTTGCTCCACAGTACAGCACAATGGCATGT 555
Qy      121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLeuGlnArgAspLeuPro 140
      556 CATCTTGTGAAGAAAGAACACATCGGCTATTCTGTTTGTGAAGCAGAGACTTTGTAACCT 615
Db      141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
      616 CAAATATATGCAAGTCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 675
Qy      161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProAlaGlu 180
      676 GCTCTGGAATTTTAAACAAACCAACACAGTGCATTTGGTGTCTCTCCACAGACTA 735
Db      181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGlnArgLeuArgAlaGlyLeu 200
      736 TGCACGTGAATAGGCAATTAGATGATGAAAGTGAATGAAGATCTAGCTGCTGCTGCT 795
Qy      201 GlyIleLeuHisAspIleGlnGlyIleArgTyrGlu-ProLysCysPro-LeuGlyValAla 220
      796 GGCATTTTATGATGATTAAGAGGATCCGCTATGAACCAATAGTCTCTCTTGAGAGTAG 855
Db      220 spIleSerLys 223
      856 ACATATCAAAA 866

```

```

RESULT 2      BQ636028      640 bp      mRNA      linear      EST 15-JUL-2002
LOCUS      h03d11.y1 Human Retina cDNA (Un-normalized, unambigified): hd/he
DEFINITION      Homo sapiens cDNA clone h03d11 5', mRNA sequence.
ACCESSION      BQ636028

```

## VERSION BQ636028.1 GI:21760487

## KEYWORDS EST.

Source Homo sapiens (human)

## ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

## REFERENCE

## AUTHORS

## TITILE

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT

## CONTACT

## SECTION

## NATIONAL

## EYE

## INSTITUTE

## 6/331, NIH, Bethesda, MD 20892-2740, USA

## Tel: 301 402 3452

## Fax: 301 496 0078

## Email: graeme@helix.nih.gov

## Plate: 03 row: d column: 11

## Seq primer: M13RP1 reverse primer (ABI).

## Location/Qualifiers

## 1. 640

## /organism="Homo sapiens"

## /mol\_type="mRNA"

## /db\_xref="taxon:9606"

## /clone="hd03d11"

## /tissue\_type="Retina"

## /dev\_stage="Adult"

## /lab\_host="EMDHI08"

## /clone\_lib="Human Retina cDNA (Un-normalized, unambigified): hd/he"

## /note="Organ: Eye; Vector: pSFOPL1; Neural retina tissue was dissected from two 80 year old donors with no observed eye disease. 100ug of total RNA was used for library construction. A directionally cloned cDNA library in the pSFOPL1 vector (Life Technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor (5'-gactatggtcttgatgacgagcgccgcc(T)15-3'). EST analysis was performed on the unambigified library at the NIH Intramural Sequencing Center (NISC)."

## ORIGIN

## Alignment Scores:

```

Pred. No.: 5.41e-117      Length: 640
Score: 1097.00      Matches: 209
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 88.47%      Indels: 0
DB: 5      Gaps: 0

```

US-10-649-273-2\_COPY\_176\_414 (1-239) x BQ636028 (1-640)

```

Qy      31 ArgArgLeuSerLeuIleLysHisIleProGlnCysSerThrMetSerGlyGlyValAlaIle 50
      3 AGAAGACTTCTTTAATAAATCAATCAAGATGCTCCACATGAGTGTGGAAAGCCATA 62
Db      51 GlnHisLeuAlaLysGlnGlyAsnArgPheHisIlePheAspIleLysProProLeuHisHis 70
      63 GAACATTTGGCCAAACAGAGAAATAGATTTCATTTGACATCAAAACCTCCCTGCATCAT 122
Qy      71 AlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMet 90
      123 GCTAAATAATGTGATTTTCTTTTACTGGAATTCACACGTTACTGATTAATAATATATG 182

```

OY		91	LysLysGluValysGlnGlyIleGluValysGlnIleLeuSerSerLalaAspIle	110
Db		183	AAAAAGAAAAGAGAAGGTATTGAAAGGGCAATCTCTTCACAGACAGCATT	242
OY		111	AlaalarvalGlnHisThrMetAlaCysHisLeuValLysArgThriSarGalalle	130
Db		243	GCTGCACAGTACACACACAATGGCGATGTCATCTTGTAATAAAGAACATCGGGCTATT	302
OY		131	LeuPheCyAlvsGlnArgAspLeuLeuProGlnnaAnaAlaValLeuValAlaSerGly	150
Db		303	CTGTTTGTGAACCAAGACTGTGTACTCAAATAAAGCAGTACTGGTGCATCTGG	362
OY		151	GlyValAlaSerAsnPhetYrIleaArgAlaLeuGlnIleLeuthraAnalathrgin	170
Db		363	GGTGCGCAATGAACTTCTATATCCGACAGCTCTGGAATTTTAAACAACGCAACAG	422
OY		171	CysThrLeuLeuCysProProProArgLeuCyStrAspAsnGlyIleMetIlealatr	190
Db		423	TGCATTTTGTGTGTCCTCTCCACAGCATATCACCTGATAAGGCCATTATGATTCATGG	482
OY		191	AsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArg	210
Db		483	AATGGTATTGAAAGCTACAGCTGCTGGCTTGGGCATTTTCAIGACATTAAGGCAATCCC	542
OY		211	TyrGlnProLysCysProLeuGlyValAspIleSerLysGluValGlYgluaIaserrile	230
Db		543	TATGAACCAAAATGTCCTCTTGAGATAGCATATCAAAAGAAAGTTGGAAGAAGCTTCATA	602
OY		231	LysValProGlnLeuLysMetGlnulle	239
Db		603	AAAGTACCACAAATTTAAAAATGAGATA	629

LOCUS	2284 bp	mRNA	linear	HTC 03-APR-2000
DEFINITION	Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length enriched library, clone:8230219017 product:similar to PUTATIVE SIALOGLYCOPROTEASE TYPE 2 [Homo sapiens], full insert sequence.			
ACCESSION	AK045669			
VERSION	AK045669.1 GI:26337528			
KEYWORDS	HTC; CAP trapper.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1			
AUTHORS	Carninci, P. and Hayashizaki, Y.			
TITLE	High-efficiency full-length cDNA cloning			
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)			
MEDLINE	99379253			
PUBMED	10349636			
REFERENCE	2			
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)			
MEDLINE	20499374			
PUBMED	11042159			
REFERENCE	3			
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, K., Yumikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.			
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer			
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)			
MEDLINE	20530913			
PUBMED	11076661			

REFERENCE	4	THE RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
AUTHORS		
TITLE		Functional annotation of a full-length mouse cDNA collection
JOURNAL		Nature 409, 685-690 (2001)
REFERENCE	5	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
AUTHORS		Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL		Nature 420, 563-573 (2002)
REFERENCE	6	(bases 1 to 2284)
AUTHORS		Adachi, J., Aizawa, K., Akiura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagasaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hoti, F., Imoto, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Nomazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaru-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission	
JOURNAL	Submitted (16-JUL-2001)	Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Chemical Research, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT		cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
		Please visit our web site for further details.
		URL: http://genome.gsc.riken.jp/
		URL: http://fantom.gsc.riken.jp/

```

FEATURES
Source

Location/Qualifiers
1..2284
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM.DB:B230219017"
/db_xref="taxon:10090"
/clone="B230219017"
/sex="male"
/tissue_type="corpora quadrigemina"
/clone_id="RIKEN full-length enriched mouse cdna library"
/dev_stage="adult"
221..1465
/note="unnamed protein product; putative
similar to PUTATIVE SIMULOGYCOPTERASE TYPE 2 [Homo
sapiens] (SPTTR|Q9HAB0, evidence: FASTY, 80%ID, 100%length,
match=1242)"
/codon_start=1
/protein_id="BAC32450.1"
/db_xref="GI:26337529"
/translation="MLMRRTAGAIKPKPKSKYGFLLRPSVARTLSGKLVLAGIET
SCDDPAAVVDERTGAVTGALSHGTOVLKMGIVPVPAGQLRRENIORIVETTLSC
RIPSDLSAATIKIPGLSLGCVGSLQVLPNPKPIPIHHHEAALATRLTNK
VEPPLVYLISGHCILALVQVSDPLILAKSLIDIAFGMDLVNARLSLIRPEEST
MSGQSLIEQLAKGGRPHFTINPMPONANKCDSPFGLQHTDKLITHEKEBGEIKG
QILSSADIAAAVQHAATCHLARTRRALIFCQKQKLLSPANAVLVSGVSNLVIR
KALEIVANATQCTLCPPRLCTDNGCIMIAMNGIERTLRGLGVLDVEDVIRREPKCPL
GVLSREVAEAAIKVPRLLKMAL"
2262..2267
/note="putative"
2284
/note="putative"

```



source

1. 1622

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="PANTOM DB:2610001M19"

/db\_xref="taxon:10090"

/clone="2610001M19"

/issue\_type="whole body"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="10 days embryo"

207..1451

/note="unnamed protein product: putative similar to PUTATIVE STALOGLYCOPROTEINASE TYPE 2 [Homo sapiens] (SPTR|Q9H4B0, evidence: FASTY, 80%ID, 100%length, match=1242)"

/codon\_start=1

/protein\_id="BAB27506.1"

/db\_xref="GI:12847276"

/translation="MLMLRTRGAIIPKPKSKVYGFRRFSVHPRTLSCHKLVIGLET  
SCDITGAADVETGVLGEALHSQTVHKTGIVPVAQOLHRENIQRIIVETLSAC  
RITPSDLAIATTIKPGIALSVGLSFLQVNOFKKPIIHMEALALIRLTNK  
VEFPVLVLISGHCIALVQSGDFLLGSKDLAPGMDLKVLRRLSLIGPST  
MSGKATIBOLAKDGRFHFTINPQNAKNCDSPFGIQTIDKLTKEKEGIEKE  
QIUSSADIAAAVQHATRCHLAKTRHAILFKQKULSPANAVLVSGVASNLTIR  
KALEIVANATQCTLCPPRLCTDNGIMIAMNGIERLRAGLVLDVEDIRYKPCPL  
GVDSIREVAEAIKVPRLKML"

1605..1610

/note="putative"

1622

/note="putative"

polya\_signal

polya\_site

ORIGIN

Alignment Scores:

Pred. No.: 2,166-113 Length: 1622

Score: 1071.00 Matches: 204

Percent Similarity: 91.21% Conservative: 14

Best Local Similarity: 85.36% Mismatches: 21

Query Match: 86.37% Indels: 0

DB: 3 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AK011265 (1-1622)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20

DB 732 CTGTGGCATATGATCCAGAGTCTTCCGATTTCCGCTCTTGGGAAGCTTTGGACANA 791

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuLysHisProGlu 40

DB 792 GCACACAGCGCAGATCTTGACACAGCTGGCAGAGACCTTTTATATCAACATCCAGAA 851

QY 41 CysSerThrMetSerGlyGlyLysValIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60

DB 852 TGTCTCAATGAGGTGGAGAAAGCTATAGAACAGTTGGCCAAAGACGGAATATGATTC 911

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80

DB 912 CATTTCATCTATCAATCCACTATGCAAGATGCTAAGAAATTCGATTTTCTTCCAGGGA 971

QY 81 LeuGlnHisValThrAspLysIleIleMetCysLysGluLysGluGluGlyIleGluLys 100

DB 972 CTTCAACATATTAATGATTAATTAACACACAGAGAAAGAAAGACGATTCGAGAG 1031

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120

DB 1032 GAGCAAAATCCGTATCATGCTGACAGATTCGCTGCGGTACACAGATCCAAACAGGTTC 1091

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnLysLeuLeuPro 140

DB 1092 CACCTTGCGAAAGAACACATGCGCTATCTGTTTGGACAGAAAGAAATTCGCTCT 1151

QY 141 GlnAsnAsnAlaValAlaLeuValAlaSerGlyValAlaSerAspPheTyrIleArgArg 160

DB 1152 CCAGCTAAGCAGATTAATGTTGATCTGAGGTGTTGCAAGTAACTGTGATCCGAAA 1211

QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180

DB 1212 GCATTGGAAATTTGCCAAATGCAACGACGATGACGTTGTGTGTCACCTCCAGACTG 1271

QY 181 CysThrAspAsnGlyIleMetIleAlaATTPAsnGlyIleGluArgLeuAsnAlaGlyLeu 200

DB 1272 TGCACTGACAAATGCAATCATGATTCATGATGGAATGGAATTGAAGATTAGCTGGCTTG 1331

QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220

DB 1332 GCGCTTTTACATATGATGATGAGACATCCGATTAAGAACCAAAATGCTCTTGGAGTAC 1391

QY 221 IleSerLysGluValAlaGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239

DB 1392 ATATCCAGAGAAAGTTCAGAAAGCTGCATTAAGTACCGGATTAATAAATGGACCTT 1448

RESULT 5

BX391919/c 852 bp mRNA linear EST 28-APR-2004

LOCUS BX391919 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens

DEFINITION CDNA clone CS0DK001YB02 3-PRIME, mRNA sequence.

ACCESSION BX391919

VERSION BX391919.2 GI:46846154

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 852)

Li,W.B., Gruber,C., Jessee,J. and Polyes,D.

Full-length CDNA libraries and normalization

Unpublished (2001)

On May 13, 2003 this sequence version replaced gi:30611736.

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 1240.r

For more information about this cluster, see [http://www.genoscope.cns.fr/cdna?c=CS0BA10352E07\\_CS03317\\_1&c=1240.r](http://www.genoscope.cns.fr/cdna?c=CS0BA10352E07_CS03317_1&c=1240.r)

FEATURES

source

1. 852

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DK001YB02"

/cell\_line="HELA CELLS COT 25-NORMALIZED"

/clone\_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.: 2,286-97 Length: 852

Score: 930.00 Matches: 180

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 75.00% Indels: 0

DB: 5 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x BX391919 (1-852)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20



Db	542	CTGTTGGCATTAAGTTCAAGAGATTCAAGTTTCGCTTCTGGAAAGCTTTGGACATA	483
Qy	21	AlaProGlyaSerMetLeuAspLysValAlaArgArgLeuSerLeuIleYshIleProGlu	40
Db	482	GCACCAAGGTGACATGCTTGACACAGGTGGCAAMACATCTTCTTAATATAAACATCCAGAG	423
Qy	41	CysSerThrMetSerGlyGlyLysAlaIleGluIleIleGluIleAlaIleGluIleAsnArgPhe	60
Db	422	TGCTTCACCAATGAGTGTGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATATATTT	363
Qy	61	HisPheAspIleYsProProLeuHisHisAlaIleYsAsnCyAspPheSerPheThrGly	80
Db	362	CATTTTGCATCAACACCTCCCTTGATCATGCTAAATAATGTGATTTTCTTTTACTGGA	303
Qy	81	LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluIleGluLys	100
Db	302	CTTCAACACGTACATGATATAATATATGAAAAAGAAAAAGAGAGATTTTGGAAG	243
Qy	101	GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys	120
Db	242	GGGCAAAATCTGTCTTCCAGCAGCAAGCATTTGCTCCACAGTACAGCACCAATGGCATTT	183
Qy	121	HisIleuValIleYsArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro	140
Db	182	CATCTGTGAAAAAGAACACATCGGGCTATTCGTTTGTGAACAGAGACATTTGTTACTT	123
Qy	141	GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg	160
Db	122	CAAAATATATGACGTACTGTTGCATCTGGTGTGTGTGCAGATTAATCTTATATCCGCAAG	63
Qy	161	AlaIleuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu	180
Db	62	GCTCTGGAATTTTAAACAAGCAACACAGTCCATTTGTGTGTCTCTCTCCAGACTA	3

REMARK	TITLE
<p>NIH-MGC project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a></p> <p>Contact: MGC help desk</p> <p>Email: <a href="mailto:cgabs-remail.nih.gov">cgabs-remail.nih.gov</a></p> <p>Tissue Procurement: Marcello Bento Soares, Ph.D.</p> <p>cDNA Library Preparation: M. Bento Soares, University of Iowa</p> <p>cDNA Library Arrayed by: The I.M.A.G.E. Consortium (IuNL)</p> <p>DNA Sequencing by: Institute for Systems Biology</p> <p><a href="http://www.systemsbio.org">http://www.systemsbio.org</a></p> <p>contact: <a href="mailto:amadamsystemsbiology.org">amadamsystemsbiology.org</a></p> <p>Anup Madan, Jessica Pahey, Erin Helton, Mark Ketterman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whitting</p>	<p>JOURNAL</p> <p>Direct Submission</p> <p>Submitted (20-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA</p>
<p>FEATURES</p> <p>source</p> <p>1. 1609</p> <p>/organism="Mus musculus"</p> <p>/mol_type="mRNA"</p> <p>/strain="C57BL/6J"</p> <p>/db_xref="taxon:10090"</p> <p>/clone="IMAGE:1226118"</p> <p>/tissue_type="Thymus gland, mouse"</p> <p>/clone_id="Soares_thymus_2NBWT"</p> <p>/lab_host="DH10B"</p> <p>/note="Vector: pT7T3-Pac"</p>	<p>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/IuNL at: <a href="http://image.lnl.gov">http://image.lnl.gov</a></p> <p>Series: IRAX plate: 66 Row: 6 Column: 10</p> <p>This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA g1: 21312463</p> <p>This clone has the following problem: frame shifted.</p> <p>Location/Qualifiers</p>

RESULT 6	BC030671	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
	BC030671		1609 bp mRNA linear HMC 19-NOV-200							
			Mus musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA							
			clone IMAGE:1226118), containing frame-shift errors.							
	BC030671			BC030671						
			GI:21040459							
			HTC.							
			Mus musculus (house mouse)							
			Mus musculus							
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
			Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.							
			1 (bases 1 to 1609)							
			Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,							
			Klausner,R.D., Collins,F.S., Wagner,L., Shemmen,C.M., Schley,G.D.							
			Alesch,I.S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,							
			Hopkins,R.F., Jordan,H., Moore,T., Wax,S.I., Wang,D., Heich,P.,							
			Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,							
			Stapleton,C., Soares,M.B., Bonaldi,M.F., Casavant,T.S.,							
			Scheer,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,							
			Carrincci,P., Prange,C., Raha,S.S., Locuelli,N.A., Peters,G.J.,							
			Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,							
			McGernan,K.J., Malek,J.A., Gunatirne,P.H., Richards,S.,							
			Wotley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Huijck,S.W.,							
			Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,							
			Fahney,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,							
			Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,							
			Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,							
			Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,							
			Butterfield,Y.S., Krzyzinski,M.I., Skalska,U., Smallus,D.E.,							
			Schermer,A., Schein,J.E., Jones,S.J., and Marra,M.A.							
			Generation and initial analysis of more than 15,000 full-length							
			human and mouse cDNA sequences							
			Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)							
			12288257							
			12477932							
			2 (bases 1 to 1609)							
			Strausberg,R.							

Alignment Scores:			
Pred. No.:	1,65e-96	Length:	1609
Score:	926.00	Matches:	183
Percent Similarity:	81.59%	Conservative:	12
Best Local Similarity:	76.57%	Mismatches:	20
Query Match:	74.68%	Indels:	24
DB:	3	Gaps:	1
US-10-649-273-2_COPY_176_414 (1-239) x BC030671 (1-1609)			
Qy 1	LeuLeuAlaLeuValAlGInGlyValSerAspPheLeuLeuLeuGlyLySerLeuAspIle	20	
Db 749	CTGTGTGGCATTAAGTCACAAAGTGTTCCGATTTCCGTCTCTTGGAAGTCTTTGACACTA	808	
Qy 21	AlaProGlyAspMetCLeuAspIysValAlaArgArgLeuSerLeuIleYsHisProGlu	40	
Db 809	GCACCAAGGCGACATGCTTGACAAAGGTGGCAAGAAAGACTTTCTTTAAATCAACATCCAGA	868	
Qy 41	CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaYsGInGlyAsnArgPhe	60	
Db 869	TGTTCTCAATGAGTGGTGGAAAAGCTATAGAACAGTTGGCCAAAGACGGAATAGATTC	928	
Qy 61	HisPheAspIleYsProProLeuHisHisAlaYsAsnYsAspPheSerPheThrGly	80	
Db 929	CATTTTACTATCAATCCACTATACAGATGCTAAGAAATTCGATTTTCTTTCACGGGA	988	
Qy 81	LeuGlnHisValThrAspIysIleIleMetYsIysGInYsGInGlyIleGluYds	100	
Db 989	CTTCAACATATTACTGATTAAGCTAATTAACACACAGGAAAAAGAAAGCATTGAGAAAG	1048	
Qy 101	GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaYs	120	
Db 1049	GGGCAAAATCCGTGATCAGCTGCAGACACTTCTCTCGCTGGCTACAGCATGCAACACGGTGC	1108	
Qy 121	HisLeuValYsArgThrHisArgAlaIleLeuPheCysYsGInArgAspLeuLeuPro	140	
Db 1109	CACCTTGGAAGAAACACATGCGGCTAATCTGTTTTGCAAGCGAAGAAATTTGCTCTTC	1168	



/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/tissue\_type="pooled"  
/lab\_host="RD10B"  
/clone\_lib="MARC 3P1C"  
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;  
library made with RNA pooled from multiple tissues  
including brain, liver, muscle, placenta/endometrium,  
ovary, testes, and bone marrow."

ORIGIN

Alignment Scores:

Pred. No.: 3.07e-91 Length: 658  
Score: 876.00 Matches: 165  
Percent Similarity: 93.68% Conservative: 13  
Best Local Similarity: 86.84% Mismatches: 12  
Query Match: 70.65% Indels: 0  
DB: Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x CP362328 (1-658)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 630 CTTTGGCATTAGTATGAGAGAGTTTCAGATTTCCTTGGACAGCTTTGACATA 571  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
DB 570 GCACCAAGGTGACATGCTTGACAGGTAGACAGACATTTCTTAATAAACAATCCACAG 511  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60  
DB 510 TGCTCCACACAGAGCGTGGAGAGCCATAGAACATTTGTCCAAACAGGAAATAAGTGG 451  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
DB 450 CATTTCGATTTCAAACCTCCCATGACAGCTGATAAAATTTGATTTTCTTTCTGCA 391  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlnGlyLysGlnGlyIleGluLys 100  
DB 390 CTTCAACATGTTATGATTAAGACAAATATGACAGAAAGAAAGAGAGATTTGAGAAG 331  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120  
DB 330 GGGCAAACTCTCTCTTCAGCTGCACAGACTGCTGTCGACGTCACAGACACATAGCTTC 271  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
DB 270 CATATTGCAAAAAGAACTCATGCTCTATTCTGTTTGCACAGAGACTTATATATGT 211  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArg 160  
DB 210 CAAGATATGCAAGTATGTTGTTATCTGAGAGTGTGCAAGTAACTTAATATACAAAA 151  
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
DB 150 GCTTAAAGATTGTGACAAATGCAACACATGCACTTGTGTCTCTCCACAGACTA 91  
QY 181 CysThrAspAsnGlyIleMetIleAlaTrp 190  
DB 90 TGCACCTGATTAATGTCATTATGATTCATGG 61

RESULT 9

CK941819/c 637 bp mRNA linear EST 15-MAR-2004

LOCUS 4065407 BARC 10BOV Bos taurus cDNA clone 10BOV12\_P24 3', mRNA

DEFINITION

ACCESSION CK941819

VERSION CK941819.1 GI:45456199

KEYWORDS EST.

SOURCE Bos taurus (cow)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euteria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.  
1 (bases 1 to 637)

REFERENCE Sonstegard, T.S., Van Tassell, C.P., Matukumalli, L.K., Harhay,  
G.P., Bosak, S., Rubenfield, M. and Gasbarre, L.C.  
Production of EST from cDNA libraries derived from immunologically  
activated bovine gut

TITLE

JOURNAL

Unpublished (2004)

Contact: Tad S. Sonstegard  
Bovine Functional Genomics Laboratory  
Animal and Natural Resources Institute  
Bldg. 200 Rm2A BARC-East, Beltsville, MD 20705, USA  
Tel: 3015048416  
Fax: 3015048414  
Email: tads@anri.barc.usda.gov

Single pass sequencing. Bases called and trimmed with phred  
0.000925 using options -trim\_at "-" -trim faasta. Vector identified  
by cross\_match using options -mismatch 12 -minscore 18  
Plate: 12 row: F column: 24  
Seq primer: AGCGATTAACAATTCACACAGC  
High quality sequence stop: 637.

FEATURES

source

1..637

/organism="Bos taurus"

/mol\_type="mRNA"

/strain="Holstein"

/db\_xref="taxon:9913"

/clone="10BOV12\_P24"

/sex="Male"

/tissue\_type="Pooled"

/dev\_stage="Multiple"

/lab\_host="DH10B T1 phage resistant"

/clone\_lib="BARC 10BOV"

/note="Organ: Small Intestine; Vector: pagen-1; Site 1:  
EcoRV; Site 2: NotI; Equimolar amounts of mRNA extracted  
from proximal jejunums of 18 and 21 wk old steers, and  
distal ileums of 14 day old calves. proximal jejunum  
exposed to C. oncophora for 3 and 6 weeks, and distal  
ileum exposed to C. parvum for 7 days"

ORIGIN

Alignment Scores:

Pred. No.: 7.81e-87 Length: 637  
Score: 838.00 Matches: 153  
Percent Similarity: 92.47% Conservative: 19  
Best Local Similarity: 82.26% Mismatches: 14  
Query Match: 67.58% Indels: 0  
DB: Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x CK941819 (1-637)

QY 54 AlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsn 73  
DB 632 GCCAAACAGGAAATGATTTGATTTGATTTCCAGCTCCATGCAACGTGCTAAAT 573  
QY 74 CysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysGln 93  
DB 572 TGTGATTTTCTTTTCTGACTTCAACAGCTTAATGATTAAGTAAATGCAAAAGAA 513  
QY 94 LysGlnGlnGlyIleGlnLysGlnIleLeuSerSerAlaAlaAspIleAlaIleThr 113  
DB 512 AAAGAGAAAGTATGACAGAGGGGAGGTCTCTTTCACCTGCAGACATTCGTGGC 453  
QY 114 ValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCys 133  
DB 452 GTGCACACACCCGTGGCTGCCCATTTGCCAAAAGAACACATGCTGCTGTGTGTC 393  
QY 134 LysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAla 153  
DB 392 AACGACAGAGGCTTCTTACCTCAGAGTAAGCAGTATGATCTGAGAGGCTTGCA 333  
QY 154 SerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeu 173

Db 332 AGTAATTATATATCCGAAAGCCCTGAAATTTGTGACCAATGCAACACAGTCGACTTTG 273

Qy 174 LeuCySPProProArgLeuCyEthrAaspAengIyIleMerIlealtrPaengIyIle 193

Db 272 CTGTGCCCCCCCCCGAGCTTTGCACTGCAACGCGGTATGATTCATGCAATGCGT 213

Qy 194 GUATGLeuAArgAlGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrrGluPro 213

Db 212 GAAAGACTAGCTGCTGGCTTGGCATTTTACACACAGAAAGCATCCGCTACGAACCA 153

Qy 214 LysCySPProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysValPro 233

Db 152 AAATGTCCTTGGAGTATATATATCAAAAGAACTTGAGAACTGCTATAAAGTCCA 93

Qy 234 GlnLeuLysMetGluIle 239

Db 92 AGATTAAAAATGAGATT 75

RESULT 10

BO961028 922 bp mRNA linear EST 21-AUG-2002

LOCUS AGENCOURT 8863711 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:6423902

DEFINITION 5', mRNA sequence.

ACCESSION BO961028

VERSION BO961028.1 GI:22376506

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 922)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: DCTD/DRP/Gazdar  
cDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Plate: LNCM2603 row: d column: 15  
High quality sequence stop: 584.

FEATURES

source location/Qualifiers

1..922

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6423902"

/tissue\_type="large cell carcinoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_1lb="NIH\_MGC\_18"

/note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; XhoI sites using the following 5' adaptor: GGCAAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC library."

ORIGIN

Alignment Scores:

Pred. No.: 4.6e-82 Length: 922

Score: 799.00 Matches: 179

Percent Similarity: 88.04% Conservative: 5

Best Local Similarity: 85.65% Mismatches: 16

Query Match: 54.44% Indels: 10

DB: 5 Gaps: 3

US-10-649-273-2\_copy\_176\_414 (1-239) x BO961028 (1-922)

Qy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20

Db 291 CTGTGGCATTAAGTTCAGAGAGTTTCAGATTTTCTGCTCTTGGAAGATCTTGGACATA 350

Qy 21 AlaProGlyLysPheLeuAspLysValAlaArgArgLeuSerLeuIleLysProGlu 40

Db 351 GCACCAAGTACATGCTTGTGCAAGTGGCAAGAAAGATCTTTTAAATAAACATCCAGAG 410

Qy 41 Cys-SerTherMetSerGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60

Db 411 TGCNTCCACCATAGTGTGGGAAAGCCATAGAACATTGCGCCAAACAGAAATTAAT 470

Qy 60 eHisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrG 80

Db 471 TCATTTGACATCAAACTCCCTTGATCATGCTAAATAATGATTTTCTTTTACTCG 530

Qy 80 yLeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGlu 100

Db 531 ACTTCAACACGTTTACTGATTAATTAATGAAGAAAGAAAGAGAGATTTTGAGAA 590

Qy 100 gGlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetLac 120

Db 591 AGGCAAAATCTGTCTTTCAGACAGACATTGCTGCCACGTAACACAAATGCGATG 650

Qy 120 sHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuP 140

Db 651 TCATCTTGTGAAAAGAA-CATCGGCTATTCGTTTGTACAGAGACATTTGATACC 709

Qy 140 OGlnAsnAsnAlaValLeuValAlaSer-GlyGlyValAlaSerAsnPheTyrlaArg 160

Db 710 TCAAAATTAATGACGTACTGCTGCTGCTGCGGGGCGCAAGTAATCTTATATCCCG 769

Qy 160 rGlnAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCySPProProAr 179

Db 770 AAACCTCGAATAATTTACAAACCCAAACCGAGGACCTTTGTGGCTCCCTCCCA 829

Qy 179 gLeuCySPThrAspAengIyIleMetIlealtrP-----AengIyIleGluArg---L 196

Db 830 ACTATGACATGATTAATGG---CATTTATGATGCTGCGGGAAGGAAATTTGAAAAATA 886

Qy 196 eutArgAlaGlyLeuGly 201

Db 887 CCNTGCTCGGCTTGGG 903

RESULT 11

AV602901/c 545 bp mRNA linear EST 27-NOV-2001

LOCUS AV602901 Bos taurus kidney fetus Bos taurus cDNA clone EIK1013A07

DEFINITION 3', mRNA sequence.

ACCESSION AV602901

VERSION AV602901.1 GI:9725227

KEYWORDS EST.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

REFERENCE 1 (bases 1 to 545)

AUTHORS Takasuga,A., Hirotsune,S., Itoh,R., Itohono,A., Suzuki,H., Aso,H.

TITLE Establishment of a high throughput EST sequencing system using poly(A) tail-removed cDNA libraries and determination of 36,000 bovine ESTs

JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)

MEDLINE 21570554

PUBMED 11713328

COMMENT Contact: Yoshihazu Sugimoto  
Animal Genetics Division  
Shirakawa Institute of Animal Genetics  
Odakura, Nishigo, Nihei-shirakawa, Fukushima 961-8061, Japan  
Tel: 81-248-25-5641  
Fax: 81-248-25-5725

Email: kazusugi@cocoa.ocn.ne.jp  
Single pass sequencing.  
This clone was obtained from a polyA-deleted cDNA library.

## FEATURES

SOURCE

Location/Qualifiers

1..545

/organism="Bos taurus"

/mol\_type="mRNA"

/db\_xref="taxon:9913"

/clone="EIKI013A07"

/tissue\_type="kidney"

/dev\_stage="fetus"

/lab\_host="DH10B"

/note="Vector: pZ19; Site 1: SalI; Site 2: NotI; Poly A was deleted from a NotI site"

## ORIGIN

## Alignment Scores:

Pred. No.:	3,186-81	Length:	545
Score:	789.00	Matches:	147
Percent Similarity:	90.11%	Conservative:	17
Best Local Similarity:	80.77%	Mismatches:	17
Query Match:	63.63%	Indels:	1
DB:	1	Gaps:	0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AV602901 (1-545)

```
OY 57 G1YAAAGPHEHISPHASPILEYSPROFROLEHISHIS-ALALYASNCYASAPRH 76
DB 544 GGAATAGATTCATTTGATTTCCANCMTCGACCAACGGTGTAAATTTGATTT 485
OY 76 ESErPHERG1YLEUG1NHISVAL1NHRASPLYS1LE1EMETLYSG1ULYSG1UG1 96
DB 484 TTCTTTTTCGACCTTCAACGCTTATTAATGATTAATGCAAGAGAAAAGAG 425
OY 96 UG1Y1LEGLYSG1ULYSG1ULYSG1ULYSG1ULYSG1ULYSG1ULYSG1ULYSG1 116
DB 424 AGGTATTCAGACGAGGCGAGGCTCTGCTTCACTGCGACATTCGTTGCGGTCCAGCA 365
OY 116 STHMETALCYSH1ELEUVALLYKATGTHHISATG1A1LELEUPHECYLSGLNAR 136
DB 364 CACCGTGGCTCCCATTCGCAAAAAGAAACACATCTGCTCTTCTTCTGCAAGCAGAG 305
OY 136 GASPLEUPROG1NANASNA1AVALLEUVAL1ASERGLYVAL1ASERASNP 156
DB 304 AGGCTTCTTACATCGAGATGACAGTACTGCTGTTGATCTGAGGCGTCCAGTAACTT 245
OY 156 EYR1LEATGARGALALEUG1ULYSG1ULYSG1ULYSG1ULYSG1ULYSG1ULYSG1 176
DB 244 ATATATTCGAAAAGCCCTGGAATTTGTGACCAATGCAACAGTGCATTTGCTGTGCC 185
OY 176 OPROPARGLYSGYETHRASPSNG1Y1LEMET1EALATPASNGLY1LEGLUARGLE 196
DB 184 GCCCCCACTCTGACATGACACACGCGCTNATGATTCGATGAAATGTTGAAAGACT 125
OY 196 UARGALAG1LYSG1ULYSG1ULYSG1ULYSG1ULYSG1ULYSG1ULYSG1ULYSG1 216
DB 124 AGGTGCTGCTTGGGCAATTTTACACAAACAGAAAGCATTCGCTACGAACCAAAATGTC 65
OY 216 OLEUG1VALASPI1SER1YSGLYVAL1GLY1UALASER1ILEYVAL1PROG1NLEU 236
DB 64 TCTTGAGATGATATATCAAAAAGATTTGAGAAAGCTGCTATTAAGTCCCAAGATTAA 5
OY 236 SNET 237
DB 4 AATG 1
RESULT 12
LOCUS CB852881 490 bp mRNA linear EST 22-APR-2003
DEFINITION UI-CF-FNO-af0-b-03-0-UI_s1 UI-CF-FNO Homo sapiens cDNA clone
ACCESSION CB852881 UI-CF-FNO-af0-b-03-0-UI 3', mRNA sequence.
```

VERSION CB852881.1 GI:30047942  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 490)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PUBMED 8889548  
COMMENT Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com) or from Open Biosystems (www.openbiosystems.com).  
Seq primer: M13 FORWARD  
POLYA=No.  
Location/Qualifiers

## FEATURES

SOURCE

1..490

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="UI-CF-FNO-af0-b-03-0-UI"

/tissue\_type="Human Lung Epithelial cells"

/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone\_lib="UI-CF-FNO"

/note="Organ: Lung; Vector: pT773-Pac (Pharmacia) with a modified polylinker; Site 1: BcoR I; Site 2: Not I; UI-CF-FNO is a subtracted cDNA library derived from two

normalized Human lung epithelial cell libraries (BN1 and DU1) The library was subtracted according to

Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. For additional information, contact:

bento-soares@uiowa.edu

TAG\_SEQ=None found"

## ORIGIN

## Alignment Scores:

Pred. No.:	1,056-80	Length:	490
Score:	784.00	Matches:	148
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	63.23%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2\_COPY\_176\_414 (1-239) x CB852881 (1-490)

```
OY 69 H1SH1S1ALYASANCYASAPHESErPHERG1YLEUG1NHISVAL1NHRASPLYS1LE 88
DB 40 CATCATGCTTAAATTTGATTTTCTTTTACGAGACTTCAACACGTTACTGATTAATA 99
OY 89 T1EMETLYSG1ULYSG1ULYSG1ULYSG1ULYSG1ULYSG1ULYSG1ULYSG1 108
DB 100 ATATGAAAAGGAAAAGGAAAGATTTGAGAAAGGCGCAAAATCTTCAAGACGA 159
OY 109 ASPI1EAL1A1NHRASPLYS1LEUG1NHISVAL1NHRASPLYS1LEUG1NHISVAL1 128
DB 160 GACATTTGCTCCACAGACACACACATGATCTTGTGAAAAGACATCGG 219
OY 129 A1A1LEUPHECYLSGLNARGASPLEUPROG1NANASNA1AVALLEUVAL1A 148
```





```
/dev_stage="36"  
/lab_host="DH10B"  
/clone_lib="CSEQC120"  
/note="Organ: limbs; Vector: pBluescript II KS(+); Site:1:  
EcoRI; Site:2: NotI; Modification of pBluescript II KS(+):  
[Stratagene] vector to accommodate cDNA produced with the  
T-primed protocol (construction of uni-directionally  
cloned cDNA libraries from messenger RNA for improved 3'  
end DNA sequencing by Glenn Fu, et al. U.S. Patent #  
6,387,624). Cut pBluescript II KS(+) with NotI and  
ligate in double stranded adaptor containing BspI and  
BamHI sites [5'ggccggcgagccgcgcgacgcaaaaaag]  
[5'aattcttttcggatccgggctcgacgc]"
```

## ORIGIN

Alignment Scores:  
Pred. No.: 9.58e-77 Length: 866  
Score: 753.00 Matches: 149  
Percent Similarity: 78.81% Conservative: 37  
Best Local Similarity: 63.14% Mismatches: 50  
Query Match: 60.73% Indels: 2  
Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x BU127463 (1-866)

```
1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlySerLeuAspIle 20  
:::|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
25 ATCTTGCGACGATGACGACGAGCTTCAGATTTCTTCTGTCGACAGTCCTAGTATA 84  
21 AlaProGlyAspMetLeuAspIleValAlaArgArgSerLeuIleLeuHisProGlu 40  
85 GCACGAGGTGACATCTGATGATAGGTAGACAGAGCTCTTTAGTGAAGCACCCGGAG 144  
41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaGlnIleValAsnArgPhe 60  
145 TGCCACGCGCATGGCTGGGGGAGGCAATAGACACCTGGCTCAACCCGAGACCTGGCA 204  
61 HisPheAspIleLeuProPheLeuHisAlaValAsnCysAspPheSerPheThrGly 80  
205 CAGTACACTTTCAGCTTCCATCCACAGCAACGATTCGTAAGTATTTCTTCTCCGGA 264  
81 LeuGlnHisValThrAspIleIleIleMetCysGlyLeuGluGluGlyIleGluIys 100  
265 CTTGACGACCTTGTGACAAAGCCATTCTTCAGAAAGAAAGAAAGATTCAGAA 324  
101 GlyIleIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120  
325 GGGGAAATCTGTCCTCGCTTAAGACATCGCTGCTGTCACAGCATGATGAGCTGCT 384  
121 HisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuLeuPro 140  
385 CATATTTATCCAGCGGACACACCGACCATGCTCTTCTTCAGTGAAGAAAACGACATATTATA 444  
141 GlnAsnAlaValIleValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160  
445 CCAAAAACCTGCAACTCTGTTCTATCAGAGAGAGTTGCAAGTATACGATATACAGAAAA 504  
161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGlu 180  
505 GAGCTGACAGACTCTGGC-AATGCAAAAGGTTTGTCTTTCTGCTCTCTCCATCAAGGCTG 563  
181 CysThrAspArgGlyIleMetIleAlaIleArgGlyIleGluArgLeuArgAlaGlyLeu 200  
564 TGCACCGAATAGGTATGATTCATGCAATGGAATGGATTGAAAGGTTCGTGACGAGATGT 623  
201 GlyIleLeuHisAspIleGluGlyIleArgGlyGluProIysCysProLeuGlyValAsp 220  
624 GGTATATA-TACAGTACGATGCGCATCCGCTATGAAACAAAGCTCCCTTGGAATTGAT 682  
221 IleSerIysGluValGlyGluAlaSerIleIysValProGlnLeuIys 236  
683 ATTTCAAAAGAGTTGAAGAGATTCCATCAGAGTCCCAAGACTAAGG 730
```

## RESULT 15

CK365185

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

CK365185 736 bp mRNA linear EST 23-DEC-2003  
AGENCOURT 17157291 NIH\_MGC 231 Rattus norvegicus cDNA clone  
IMAGE:7097620 5', mRNA sequence.  
CK365185.1 GI:40331120  
EST.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 736)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Howard Jacobs  
cDNA Library Preparation: Express Genomics  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: LHM14944 row: d column: 02  
High quality sequence stop: 736.  
Location/Qualifiers  
1..736  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
/clone="IMAGE:7097620"  
/tissue\_type="lung, pooled"  
/lab\_host="DH10B Tona"  
/clone\_lib="NIH\_MGC\_231"  
/note="Organ: lung; Vector: pExpress-1; Site:1: EcoRV/  
Site:2: NotI; RNA obtained from pooled lung tissue from a  
mix of male and female animals at 8 wk old. Tissues were  
snap-frozen and kept at -80C for two days before RNA  
extraction and purification (TRI-reagent method). cDNA was  
primed using oligo-dT primer:  
5'-pGACTGTTCTTAATGCGGAGCGCGCCGCC(T)25-3' and cloned into  
the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb  
resulted in an average insert size of 2.3 kb. This primary  
library is not normalized (normalized primary library is  
NIH\_MGC 232) and was constructed by Express Genomics  
(Frederick, MD). Note: this is a NIH\_MGC library."

## ORIGIN

Alignment Scores:

Pred. No.: 2.22e-76 Length: 736  
Score: 749.00 Matches: 141  
Percent Similarity: 91.02% Conservative: 11  
Best Local Similarity: 84.43% Mismatches: 15  
Query Match: 60.40% Indels: 0  
Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x CK365185 (1-736)

```
73 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleMetCysIys 92  
1 AAGCTGATTTTCTTTTACGGGACCTCAACATGCAACCATAGCTAATATACACACAG 60  
93 GluIysGluGluGlyIleGluIysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 112  
61 GAAAAAAGAAAGAGCATTTGAAAGGGGCAATCTCTTCATCAGCGGACGACATTTGCTCT 120  
113 ThrValGlnHisThrMetAlaCysHisLeuValIysArgThrHisArgAlaIleLeuPhe 132
```



```
Db      121 GCGGTACGACGCGCAACAGCGTGCCACCTTGCGAAAAGAACACATCGTCTATTCTGTT 180
Qy      133 CysIysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 152
Db      181 TGCCAGCAGAAATAATTGCTATCTCCAGCTAAGCAGTATTAGTTGTCCTGGAGGCTTT 240
Qy      153 AlaSerAsnPheTyrTlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 172
Db      241 GCAAGTAACTTGTAACATCCGAGAGCATTTGAAATTGTAGCAAAATGCAACACATGCACT 300
Qy      173 LeuLeuCysProProProArgLeuLeuThrAspAsnGlyIleMetIleAlaTTPaengly 192
Db      301 TTGTGTGTGTCCTCCGAGCTGTCAGCTGACAAATGATCATGATGTCATGGAATGGA 360
Qy      193 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 212
Db      361 ATTGAAGAATTACGTCGCTGGCTTTGGGCATTTTACATGATGTAGAAAGACATCCGATACGA 420
Qy      213 ProIysCysProLeuGlyValAspIleSerIysGlyValGlyGluAlaSerIleIysVal 232
Db      421 CCAAAATGTCTCTCGAATAGACATATCCAGAGAAAGTTGCAAGAGCTGCCATAAAAGTA 480
Qy      233 ProGlnLeuIysMetGluIle 239
Db      481 CCACGATTAAAAATGACACTT 501
```

Search completed: November 11, 2005, 02:14:17  
Job time : 3392.57 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: November 10, 2005, 16:54:39 ; Search time 169.638 Seconds  
(without alignments)  
2305.321 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 L1ALVGVSDFLILGKSLDI.....DISKVGESAIKVPOLKMEI 239

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=x1p  
-Q=/cgn2\_1/USFTO.spool\_P/US10649273/runat\_02112005\_091339\_15581/app\_query.fasta\_1.1429  
-DB=Issued\_Patents\_NA -OFMT=faetap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bl0sum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=plo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273 @CNC 1.1 177 @runat\_02112005\_091339\_15581 -NCPUR=6 -ICPF=3  
-NO\_MAP -LAGGEOUTERY -NEG\_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1240	100.0	1416	4	US-09-774-528-177
2	1240	100.0	1526	4	US-10-067-443-23
3	1240	100.0	2197	4	US-10-067-443-1
4	1213.5	97.9	1387	4	US-10-067-443-21
5	725	58.5	14364	4	US-10-067-443-20
6	248.5	20.0	94750	4	US-09-596-002-38
7	247	19.9	1053	4	US-09-540-236-806
8	231	18.6	1059	4	US-09-252-991A-884
9	231	18.6	1206	4	US-09-252-991A-801
10	224	18.1	1830121	4	US-09-557-884-1
11	224	18.1	1830121	4	US-09-643-990A-1
12	221.5	17.9	996	4	US-09-902-540-6612

C	13	221.5	17.9	2582	4	US-09-902-540-503	Sequence 503, App
	14	212	17.1	1074	4	US-09-543-681A-2341	Sequence 2341, Ap
	15	205	16.5	1372	4	US-09-489-033A-2050	Sequence 2050, Ap
	16	200	16.1	1315	1	US-08-087-797-1	Sequence 1, Appli
	17	197	15.9	1092	4	US-09-107-533A-2955	Sequence 2955, Ap
	18	190	15.3	1008	3	US-08-987-121A-5	Sequence 5, Appli
	19	190	15.3	1011	3	US-08-987-121A-3	Sequence 3, Appli
	20	189	15.2	822	4	US-09-710-279-727	Sequence 727, App
	21	189	15.2	3993	4	US-09-710-279-3985	Sequence 3985, Ap
	22	187	15.1	1155	4	US-09-602-777A-147	Sequence 147, App
	23	186.5	15.0	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	24	186.5	15.0	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	25	186	15.0	1006	3	US-08-961-083-51	Sequence 51, Appli
	26	186	15.0	1006	4	US-09-536-784-51	Sequence 51, Appli
	27	186	15.0	1107	3	US-09-134-001C-1072	Sequence 1072, Ap
	28	186	15.0	10974	4	US-08-961-527-214	Sequence 214, App
	29	184	14.8	1011	3	US-09-066-512-1	Sequence 1, Appli
	30	182	14.7	1011	4	US-09-583-110-2196	Sequence 2196, Ap
	31	182	14.7	1633	4	US-09-620-312D-6	Sequence 6, Appli
	32	181.5	14.6	1664976	4	US-08-328-352-261	Sequence 261, App
	33	180.5	14.6	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	34	180.5	14.6	1664976	4	US-09-692-570-1	Sequence 1, Appli
	35	178	14.4	640681	4	US-09-790-988-1	Sequence 1, Appli
	36	177	14.3	1011	4	US-09-107-433-1618	Sequence 1618, Ap
	37	174.5	14.1	1230025	4	US-09-198-452A-1	Sequence 1, Appli
	38	174.5	14.1	1230230	4	US-09-438-185A-1	Sequence 1, Appli
	39	174	14.0	1026	3	US-09-149-624-1	Sequence 1551, Ap
	40	173	14.0	1101	4	US-08-956-171E-102	Sequence 102, App
	41	170	13.7	15249	4	US-08-781-986A-102	Sequence 102, App
	42	170	13.7	15249	4	US-08-781-986A-102	Sequence 794, App
	43	168	13.5	3064	3	US-09-221-017B-794	Sequence 131, App
	44	159.5	12.9	42325	4	US-08-311-731A-131	Sequence 3566, Ap
	45	145.5	11.7	3215	4	US-09-710-279-3566	

ALIGNMENTS

US-09-774-528-177  
Sequence 177, Application US/09774528  
Patent No. 6743619  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Zhou, Ying  
APPLICANT: Goodrich, Ryle  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Ren, Feiyang  
APPLICANT: Zhang, Jie  
APPLICANT: Zhao, Qing A.  
APPLICANT: Yang, Yonghong  
APPLICANT: Xue, Aidong J.  
APPLICANT: Wehrman, Tom  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Wang, Dunrui  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. 6743619el Nucleic Acids and  
TITLE OF INVENTION: Polypeptides  
FILE REFERENCE: 802  
CURRENT APPLICATION NUMBER: US/09/774,528  
CURRENT FILING DATE: 2001-01-30  
NUMBER OF SEQ ID NOS: 441  
SOFTWARE: PL\_FU\_genes Version 2.0  
SEQ ID NO 177  
LENGTH: 1416  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURES:  
NAME/KEY: CDS  
LOCATION: (205) .. (1305)  
US-09-774-528-177

Alignment Scores:

Pred. No.: 1,02e-157 Length: 1416  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-09-774-528-177 (1-1416)

```
QY 1 LeuenuAlaValGInGlyValSerApPheLeuLeuGlyLysSerLeuApPle 20
Db 586 CTGTTGGCATTAGTTCAAGAGGTTTCAGATTTCTGCTTCTTGAAAGCTTTGACATA 645
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuLysHisProGlu 40
Db 646 GCACCAAGTGACATGCTTGACAGAGTGCCAGAAAGACTTCTTTATATAACATCCAGAG 705
QY 41 CysSerTherMetSerGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
Db 706 TGCTCCACCATGAGTGGTGGAAGCCATAGAACATTTGGCCAAACAGAAATGATTT 765
QY 61 HisPheAspLysProProLeuHisAlaLysAsnCysAspPheSerPheTherGly 80
Db 766 CATTTCACATCAACCTCCCTGCATCATCTAAAAATGATTTTCTTTTATCTGGA 825
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnGlyLysGlnGlyLys 100
Db 826 CTTCAACACGTTACTGATTAATTAATGAAAAAGAAAAAGAGAGCTTTGACAGAG 885
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisTherMetAlaCys 120
Db 886 GGGCAAAATCCTGCTTCAGCAGCAGACATGCTGCCACAGTACACACAAATGCGCATGT 945
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
Db 946 CATCTTGAAAGAACACATCGGGCTATCTGTTTGTAAAGACAGAGACTGTTGACTT 1005
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheTyrIleArgArg 160
Db 1006 CAAATTAATGAGTACTGTTGCTGATCTGCTGCTGCGAAGTAACTTTATATCCGCGACA 1065
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db 1066 GCTCTGGAATTTTAAACAAAGCACAACAGTCACTTTGTTGTGTCTCTCCAGACATA 1125
QY 181 CysThrAspAsnGlyLysMetIleAlaIleArgArgGlyLysGlnArgLeuArgAlaGlyLeu 200
Db 1126 TGCACTGATTAATGCGATTAATGATGCAATGATGTAAGAAAGACTACGTCCTGCTTG 1185
QY 201 GlyIleLeuHisAspIleGluGlyLysIleArgTyrGlnProLysCysProLeuGlnValAsp 220
Db 1186 GGCATTTTTACATGACATAGAAAGCATCGCTATGAAACCAAAATGTCCTCTTGAGATGAC 1245
QY 221 IleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGluIle 239
Db 1246 ATATCAAAAGAGATTGGAAGACTTCATTAAGTACCAATTAATAATGAGATTA 1302

RESULT 2
US-10-067-443-23
; Sequence 23, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
```

SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 23  
; LENGTH: 1526  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-067-443-23

Alignment Scores:

Pred. No.: 1.15e-157 Length: 1526  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-067-443-23 (1-1526)

```
QY 1 LeuenuAlaValGInGlyValSerApPheLeuLeuGlyLysSerLeuApPle 20
Db 85 CTGTTGGCATTAGTTCAAGAGGTTTCAGATTTCTGCTTCTTGAAAGCTTTGACATA 144
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuLysHisProGlu 40
Db 145 GCACCAAGTGACATGCTTGACAGAGTGCCAGAAAGACTTCTTTATATAACATCCAGAG 204
QY 41 CysSerTherMetSerGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
Db 205 TGCTCCACCATGAGTGGTGGAAGCCATAGAACATTTGGCCAAACAGAAATGATTT 264
QY 61 HisPheAspLysProProLeuHisAlaLysAsnCysAspPheSerPheTherGly 80
Db 265 CATTTCACATCAACCTCCCTGCATCATCTAAAAATGATTTTCTTTTATCTGGA 324
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnGlyLysGlnGlyLys 100
Db 325 CTTCAACACGTTACTGATTAATTAATGAAAAAGAAAAAGAGAGCTTTGACAGAG 384
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisTherMetAlaCys 120
Db 385 GGGCAAAATCCTGCTTCAGCAGCAGACATGCTGCCACAGTACACACAAATGCGCATGT 444
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
Db 445 CATCTTGAAAGAACACATCGGGCTATCTGTTTGTAAAGACAGAGACTGTTGACTT 504
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheTyrIleArgArg 160
Db 505 CAAATTAATGAGTACTGTTGCTGATCTGCTGCTGCGAAGTAACTTTATATCCGCGACA 564
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db 565 GCTCTGGAATTTTAAACAAAGCACAACAGTCACTTTGTTGTGTCTCTCCAGACATA 624
QY 181 CysThrAspAsnGlyLysMetIleAlaIleArgArgGlyLysGlnArgLeuArgAlaGlyLeu 200
Db 625 TGCACTGATTAATGCGATTAATGATGCAATGATGTAAGAAAGACTACGTCCTGCTTG 684
QY 201 GlyIleLeuHisAspIleGluGlyLysIleArgTyrGlnProLysCysProLeuGlnValAsp 220
Db 685 GGCATTTTTACATGACATAGAAAGCATCGCTATGAAACCAAAATGTCCTCTTGAGATGAC 744
QY 221 IleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGluIle 239
Db 745 ATATCAAAAGAGATTGGAAGACTTCATTAAGTACCAATTAATAATGAGATTA 801

RESULT 3
US-10-067-443-1
; Sequence 1, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: SPINAL CORD, MP-1
```

```
FILE REFERENCE: D0073 NP
CURRENT APPLICATION NUMBER: US/10/067,443
CURRENT FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 2197
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (231)..(1472)
US-10-067-443-1

Alignment Scores:
Pred. No.: 2,07e-157 Length: 2197
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-1 (1-2197)
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyValSerLeuAsp11e 20
DB 756 CTGTTGGCATTAATTCAGAGAGTTTCAGATTTTCCTTTGGAAAGTCTTTGACATA 815
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 816 GCACCGAGTGACATGCTTGCAAGAGTGCAAGAGCTTTCTTAATAAATCATCCAGAG 875
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 876 TGGTCCACCATGAGTGGTGGAAGCATATGAACTTTGGCCAAACAGAAATATGATTT 935
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
DB 936 CATTTTGACATCAAACTCCCTTCATCATGCTAAATATGATTTCTTTTACTGGA 995
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100
DB 996 CTTCAACAGCTATGATTAATAATATATGAAAAAGAGAGAGATTTGAGAG 1055
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120
DB 1056 GGGCAAACTCTGCTTCAGACAGACATTCCTCCACAGTACGACACATGGCATGT 1115
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
DB 1116 CATCTTGTAAGAAAGAACACATCGGCTATTCGTTTGTAAAGCAGAGACTTGTACT 1175
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaLysSerAsnPheTyrIleArgArg 160
DB 1176 CAAATAAATGACATCTGCTGATCTGATGATGCTGCAAGTAACTTCTATATCCGAGA 1235
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
DB 1236 GCTCTGAAATTTTAACAAACGCAACAGTGCCTTTGTTGTCCTCTCCCAAGACTA 1295
QY 181 CysThrAspAsnGlyIleMetIleAlaTyrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1296 TGCACGATATATGCACTTATGATTTGCAATGAAATGATTTGAAGACATACGTCGGCTTG 1355
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
DB 1356 GGCATTTTACATGACATTAAGAGCATCGCTATTAACCAAAATGCTCTTGAGTAGAC 1415
QY 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
```

```
DB 1416 AATCAAAAAGAGTTGGAGAGCTTCATTAAGTACCAATTAATAATGAGATA 1472
RESULT 4
US-10-067-443-21
Sequence 21, Application US/10067443
Patent No. 6642041
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
TITLE OF INVENTION: SPINAL CORD, MP-1
FILE REFERENCE: D0073 NP
CURRENT APPLICATION NUMBER: US/10/067,443
PRIOR FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.0
SEQ ID NO 21
LENGTH: 1387
TYPE: DNA
ORGANISM: homo sapiens
US-10-067-443-21

Alignment Scores:
Pred. No.: 3.87e-154 Length: 1387
Score: 1213.50 Matches: 238
Percent Similarity: 90.15% Conservative: 0
Best Local Similarity: 90.15% Mismatches: 1
Query Match: 97.86% Indels: 25
DB: Gaps: 1

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-21 (1-1387)
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyValSerLeuAsp11e 20
DB 549 CTGTTGGCATTAATTCAGAGAGTTTCAGATTTTCCTTTGGAAAGTCTTTGACATA 608
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 609 GCACCGAGTGACATGCTTGCAAGAGTGCAAGAGCTTTCTTAATAAATCATCCAGAG 668
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 669 TGGTCCACCATGAGTGGTGGAAGCATATGAACTTTGGCCAAACAGAAATATGATTT 728
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
DB 729 CATTTTGACATCAAACTCCCTTCATCATGCTAAATATGATTTCTTTTACTGGA 788
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIle----- 98
DB 789 CTTCAACAGCTTATGATTAATAATATATGAAAAAGAGAGAGATTAATTTCTA 848
QY 98 ----- 98
DB 849 AATGATAAGTTGAACAGATTAATATTCCTGATTTGCTTAATAATAGCTGCTATTC 908
QY 99 -----GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGln 115
DB 909 TGCACGATATGCACTTCTGTAAGAAAGACATCGGCTATTCCTTTTGTAAAGAG 968
QY 116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135
DB 969 CACACATGCGATGATCATCTTGTAAAGAAACATCGGCTATTCCTTTTGTAAAGAG 1028
QY 136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaLysAsn 155
DB 1029 AGAGACTTGTAACTCAAAATAATATGACATGCTGATGATGCTGATGCTGCAAGTAAAC 1088
QY 156 PheTyrIleArgAlaGlnAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 175
```

```
Db 1089 TTCTAATATCCGACAGCTCTGGAATTTTAAACAACGACACAGTGCATTTGTGTGT 1148
Qy 176 ProProProArgLeuCystrHraAspAsnGlylleMetIleAaTPasGlylleGluArg 195
Db 1149 CTTCTCCACGACATACACATGATTAAGCATTAATGATTCATGATTAAGTGAAGAA 1208
Qy 196 LeuArgAlaGlyLeuGlylleLeuHraAspIleGluGlylleArgTyGluProlysCys 215
Db 1209 CTACGCGTGGCTTGGGCATTTTACATGACATAGAGGATCCGCTATGAACCAAAATGT 1268
Qy 216 ProLeuGlyValAspIleSerLySGluValGlylualaSerIleLySvalProGluLeu 235
Db 1269 CCTCTTGAGATGACATATCAAAAGAGTTGAGAGAGCTTCATTAAGATCACCAATTA 1328
Qy 236 LyMetGluIle 239
Db 1329 AAAATGAGATA 1340

RESULT 5
US-10-067-443-20
; Sequence 20, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED I
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 14364
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-20

Alignment Scores:
Pred. No.: 3 21e-86 Length: 14364
Score: 725.00 Matches: 186
Percent Similarity: 32.86% Conservative: 0
Best Local Similarity: 32.86% Mismatches: 0
Query Match: 58.47% Indels: 380
Gaps: 3

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-20 (1-14364)
Qy 29 ValAlaArgArgLeuSerLeuIleLyshisProGluCysSerThmeSerGlyLyS 48
Db 11840 GTGGCAAGAGACTTTCTTTAATAAATCAATCCAGAGTCTCCACCATAGATGTGGAAA 11899
Qy 49 AlaIleGluHisLeuAlaLySGlnGlyAsnArgPheHisPheAspIleLySProLeu 68
Db 11900 GCCATAGAACATTTGGCCAAACAGAAATAGATTTCATTTTGAATCAAACTCCCTTG 11959
Qy 69 HisHisAlaLySAsnCySAspPheSerPheThrGlyLeuGlnHisValThraSrpLyS 88
Db 11960 CATCATCTAATAAATTTGATTTTCTTTTACGTGACATCAACGTTACGTATAAATA 12019
Qy 89 IleMetLySArgLySGluGlu-----GlyIleGluLySGlnI 96
Db 12020 ATATATGAAAAAGAAAAAGAGAAAGTATATTCTAATTAGTAAGATTGAACAGATAAT 12079
Qy 97 -----GlyIleGluLySGlnI 103
Db 12080 ATTCTCGATTGTGCTTAATAAATAGTGTCTCATTTCTGACAGGTATTTGAGAGGGCAAT 12139
Qy 103 eleuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCyShisLeuVa 123
```

```
Db 12140 CCTGTCTTCAGACAGACATTTGCTGCAACATACAGCACAAATGGCATGTCATCTTGT 12199
Qy 123 IlysaArgThrHisArgAlaIleLeuPheCySArgGlnArgAspLeuProGlnAsnAs 143
Db 12200 GAAAAAGAACATCGGGCTATTCTGTTTGTATAGACAGAGACTGTCTTCAAAATTA 12259
Qy 143 naIaValIleu----- 146
Db 12260 TGCAGTACTGTAAGTTTATCTCATTTTATAGTATAGTTACATTGCAATATGTTC 12319
Qy 146 ----- 146
Db 12320 TTTTTCACAGACCTTGACCTTGATGTTAGATGAACAGATCTTATGCTATGCTAG 12379
Qy 146 ----- 146
Db 12380 CCTGACAGTATGAATAATTATGACAGATAGAAAAACTAACGCCATTCTTGATAGTT 12439
Qy 146 ----- 146
Db 12440 TGTAGCTTTATAGGACAGCTGTATAGCTTTATGSCACATAAGTCAATTTTGCATCTT 12499
Qy 146 ----- 146
Db 12500 CTGTGTGATTTAAAAAGAGGCTTACAAATTAAGAAAGTAATGACAGTAACTGCTATCACT 12559
Qy 146 ----- 146
Db 12560 ATTTTGAAGAAATAGTNGATTTCTTCATCTTGTATGAATAATCCCTTTGTTGTGT 12619
Qy 146 ----- 146
Db 12620 TTTTAAATAGACAGCTCAATTTTACAGTGGAGGTGTATTCACACTTCGTGACACT 12679
Qy 146 ----- 146
Db 12680 AATGTGATTAAGTTCTGATTAATCCATATATTTGACAGCAAAATCCCTTAATGTG 12739
Qy 146 ----- 146
Db 12740 CTTAAAGCCTTGACAAACATCTGTTTAACTGTATCTTAACTTTATCATTTAAAT 12799
Qy 146 ----- 146
Db 12800 TATTAATCTAAGTGGAAAAATGTTTAAATGTAAGTATTCATATGATGAATTTACATG 12859
Qy 146 ----- 146
Db 12860 ATATCAAAAGATTAATTTTTCAGAGTTATGTAGTAATAATGCACAAATAATAAAATTC 12919
Qy 146 ----- 146
Db 12920 AGGGTCTAATAATAGTGTACTAGATTAATTAATTAATTAATTAATTAATTAATTAAT 12979
Qy 146 ----- 146
Db 12980 TGAAGAAAAATATACAAAAATGCTAGTAATGTTTGTATGCTATTAAGATTAATTA 13039
Qy 146 ----- 146
Db 13040 TTTTCTTCCAAATTTTATTAATAGATATGTCATGTCGCCATTACCATGTCAAAT 13099
Qy 146 ----- 146
Db 13100 GGAATGTTTATATGTTAATGCTATATTTTCTCCAGTTTATTAAGACTTGTGT 13159
Qy 147 -----ValAlaSerGly 151
Db 13160 TCATATCCATATATGATAGTATTTTGTGTTTCTCAATCTTCAGGTGCACTGTGTG 13219
Qy 151 yValaIaSerAsnPheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCy 171
```

```
Db 13220 TGTGGCAGTAATTCTATATCCGACAGCTCTGGAATTTTAACAACGACACAGTCG 13279
QY 171 sThrLeuLeuCySPProProIaArgLeuCySthzAspAnglyIleMetIleAlaTrp-- 190
Db 13280 CACTTGTGTGTCTCTCTCCACAGCTATGACGTATATGSCATGATGATTGATGATGGA 13339
QY 190 ----- 190
Db 13340 AGCCACAGATATACGTGCTTCACTCATATCTATGTAATTAATTCGCATTTATCAT 13399
QY 190 ----- 190
Db 13400 ACTAGCCTTCTCTCTTCCAGATCTTGAGCTATGATTTATTTAAATGCTTCTTATTTA 13459
QY 191 --AenglyIleGlyArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlyIleA 210
Db 13460 GGAATGGATATTGAAAGATACGTGCTGCTTGGCATTTTACATGACATGAGAAGCATCC 13519
QY 210 rGlyrGlyuProLys 214
Db 13520 GCTATGAACCAAG 13533

RESULT 6
US-09-596-002-38
; Sequence 38, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Paterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 38
; LENGTH: 94750
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte template ID No. 6632636 38
; PUBLICATION INFORMATION:
; US-09-596-002-38

Alignment Scores:
Pred. No.: 3,02e-20 Length: 94750
Score: 248.50 Matches: 73
Percent Similarity: 48.67% Conservative: 37
Best Local Similarity: 32.30% Mismatches: 90
Query Match: 20.04% Indels: 27
Gaps: 7

US-10-649-273-2_COPY_176_414 (1-239) x US-09-596-002-38 (1-94750)
QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db 24305 ATGCTGTGGCGTGCATGCTGGCGCTGATCAGATATTCGGGAGCTTATCATGAT 24364
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db 24365 GCGGTGGGTGAATCTTGTGATTAACGCGCAAAATGCTCAAACTG---CCCTATCT--- 24418
QY 41 CysSerThrMetSerGlyLysAlaIleGluHisLeuAlaLysGInGlyAspArgPhe 60
Db 24419 -----GGTGGCCCAAAATTCGAAAAATTAAGCAAAAACGCAACCCACAC 24463
QY 61 HisPheAspIleLysProProLeuHisIleAlaLysAsnCyAspPheSerPheThrGly 80
Db 24464 GCGTATAGCTGCCAAGACCAATGACAGAT---AAAGGCTGAGATTTTTCGTCAGTGCC 24520
```

```
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGlyLys 100
Db 24521 ATGAAACCGCATTCATATCATCAAGACACCAACGCAAGCCAGCCACCC--- 24577
QY 101 GlyIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisIleMetAlaCys 120
Db 24578 -----GCCACAGACGACATCGCGCAAGCTTTGATGATGGCGGTGGAT 24625
QY 121 HisLeuValLysArgThrHisIleArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
Db 24626 ACTTGGTCAAAAATGACCAAGCACTACAGATACAGCATTCGCCAG----- 24676
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyLysValAlaSerAsnPheTyrlleArgArg 160
Db 24677 -----CTGGTGGTCGACGAGGGGGGTCTCTGCGCATGAGATGCACGCCCG 24721
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCySthzLeuLeuCySPProProArgLeu 180
Db 24722 ACCCTGACCGAGACGCTCCGCAAAATCGATGCTGCTGATCATGCCCCGACGACTA 24781
QY 181 CyethrAspAnglyIleMetIleAlaTrpAnglyIleGlnArgLeuArgAlaGlyLeu 200
Db 24782 TGCACGATATATGTCGATGATCGCTATGCTGCTTTGTGCTGCTGACGCTGTGACAG 24841
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrgLupProLysCySPProLeuGlyValAsp 220
Db 24842 TGC-----GATGACTTGGC-GTTTCGCTGATTC-----CCGATGGATATGAC 24885
QY 221 IleSerLysGlnValGly 226
Db 24886 GACGCTGGCGTATCGGC 24903

RESULT 7
US-09-540-236-806
; Sequence 806, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709,2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 806
; LENGTH: 1053
; TYPE: DNA
; ORGANISM: M. catarrhalis
; US-09-540-236-806

Alignment Scores:
Pred. No.: 3,22e-23 Length: 1053
Score: 247.00 Matches: 70
Percent Similarity: 46.85% Conservative: 34
Best Local Similarity: 31.53% Mismatches: 80
Query Match: 19.92% Indels: 38
Gaps: 6

US-10-649-273-2_COPY_176_414 (1-239) x US-09-540-236-806 (1-1053)
QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
Db 445 ATGCTGTGGCGTGCATGCTGGCGCTGATCAGATATTCGGGAGCTTATCATGAT 504
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
Db 505 GCGGTGGGTGAATCTTGTGATTAACGCGCAAAATGCTCAAACTG---CCCTATCT--- 558
QY 41 CysSerThrMetSerGlyLysAlaIleGluHisLeuAlaLysGInGlyAspArgPhe 60
Db 559 -----GGTGGCCCAAAATTCGAAAAATTAAGCAAAAACGCAACCCACAC 603
```





Db 604 TTGTCGGCGGTGAGGATATCGCCGCTACCACTTCTTGGCAATCGGATCGGACATGCC 663  
Qy 22 ProGlyAspMetLeuAspIleValAlaArgArgLeuSerLeuIleuYshisProGluCys 41  
Db 664 GCCGCGAAGCTTGCACAAAGACCGCAAGCTGATGGCCTG---GGTATGCC----- 714  
Qy 42 SerThiMetSerGlyGlyValAlaIleGluHisLeuAlaIleGluHisArgPheHis 61  
Db 715 -----GGTGTCCGGAATTCGCCCGCTGCGGACGCGGCACTCTGCGCCG 762  
Qy 62 PheAspIleLeuProLeuHisHisAlaIleValAsnCysAspPheSerPheThiArgIleu 81  
Db 763 TTCGTTCCCGCGCGCGCGATGACCCATCGCCCGCTGAGCTTCAAGCTTCAAGCGGCGTC 822  
Qy 82 GlnHisValThrAspIleIleMetIleGluGlyGluGluGluGluGluGluGluGlu 101  
Db 823 AAGACCTTTACCTGAC---ACCTGGCAGCGTTGGTGTGAGGCGCGGACACAGACGAG 879  
Qy 102 GlnIleLeuSerSerAlaIleAspIleAlaIleValAlaIleGluHisThiMetAlaCysHis 121  
Db 880 CAG-----ACCGCTGCACATCGCCCTGCGGCTTCCAGACCGCGGCTGTGAGACC 930  
Qy 122 LeuValIleArgThrHisArgAlaIleLeuPheCysIleGluArgAspIleuProGln 141  
Db 931 CTGCTGATCAAGTCCGCTGCGCGCTTG-----AAGCAGACCGCGCTG-----AAG 975  
Qy 142 AsnAsnAlaValIleValAlaSerGlyValAlaSerAspPheThiArgArgAla 161  
Db 976 AAC-----CTGTATGCGCGCGCGGTCTACGCGCACAGCGCGCTGCGCAGCGGC 1026  
Qy 162 LeuGluIleLeuThrAsnAlaThrGlnCysThiLeuLeuCysProProArgIleuCys 181  
Db 1027 CTGGAAGAAGATGCTCGCGCAATGAAGGGGCGAGGTCTTACGCGCGCGCGCTTCTTC 1086  
Qy 182 ThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyIleuGly 201  
Db 1087 ACCGCAATGGCGCATATCGCTACGCGCTGCGCAGCGCGCTGCTGCGCGCGC----- 1140  
Qy 202 IleuHisAspIleGluGlyIleArgIleGluProIleCysProIleu 217  
Db 1141 ---CAGCATGAAGCGCGCGCGATCAGCTCCAGCGCGCGCTGCGCGCATG 1185  
RESULT 10  
US-09-557-884-1/c  
; Sequence 1, Application US/09557884  
; Patent No. 6506581  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: The Nucleotide sequence of  
; the Haemophilus influenzae Rd Genome, Fragments  
; thereof, and Uses Thereof  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3 1/2 inch diskette  
; COMPUTER: Dell Pentium  
; OPERATING SYSTEM: MS DOS v6.22  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/557,884  
; FILING DATE: 25-Apr-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 06/476,102  
; FILING DATE: JUN-5-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks

; REGISTRATION NUMBER: 41,971  
; REFERENCE/DOCKET NUMBER: PB186P3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-309-8504  
; TELEFAX: 301-309-8439  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1830121 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-557-884-1  
Alignment Scores:  
Pred. No.: 7,8e-15 Length: 1830121  
Score: 224.00 Matches: 67  
Percent Similarity: 47.50% Conservative: 28  
Best Local Similarity: 33.50% Mismatches: 83  
Query Match: 18.06% Indels: 22  
DB: 4 Gaps: 6  
US-10-649-273-2\_COPY\_176\_414 (1-239) x US-09-557-884-1 (1-1830121)  
Qy 2 LeuAlaIleuValGlnGlyValSerAspPheLeuLeuGlyIleYshisProGluCys 21  
Db 552791 TTAGTCGTCGTGATGATGTAAGTAAGTATGAGAAATCTATGATGATGCT 552732  
Qy 22 ProGlyAspMetLeuAspIleValAlaArgArgLeuSerLeuIleuYshisProGluCys 41  
Db 552731 GCTGGCGAAGCTTTATTAACAGCAAAATTAATCTGACTA---GATTATCCA----- 552681  
Qy 42 SerThiMetSerGlyGlyValAlaIleGluHisLeuAlaIleGluHisArgPheHis 61  
Db 552680 -----GTTGCGCGCGGACTTCTCTGTTAGCGGAAAGCTACGCCCAATTCGT 552633  
Qy 62 PheAspIleLeuProLeuHisHisAlaIleValAsnCysAspPheSerPheThiArgIleu 81  
Db 552632 TTACACTTTCCACGTCATGACAGATCGTCAGCGCTGATTTAGTTTCTGCTTTA 552573  
Qy 82 GlnHisValThrAspIleIleMetIleValGluGluGluGluGluGluGluGluGlu 99  
Db 552572 AAAACATTTGCCGCAATACAGTTATCAAGCAATTAACAGAGCGGCAACTGATAGAG 552513  
Qy 100 IysGlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisThiMetAla 119  
Db 552512 CAA-----ACTAAAGCAGATATGCTTATGCTTTCACAGATGCGGTGTG 552468  
Qy 120 CysHisLeuValIleArgThrHisArgAlaIleLeuPheCysIleGluArgAspIleu 139  
Db 552467 GATACTCTTGCC-----ATTAAATGTAAAG---CGTGATTTGAAA 552432  
Qy 140 ProGlnAsnAsnAlaValIleuValAlaSerGlyValAlaSerAspPheThiArg 159  
Db 552431 GAAACAGCTAATAACGTTATGATGATGGGAGGGGTGACGCAATTAACAACTCCA 552372  
Qy 160 ArgAlaIleuGluIleuThrAsnAlaThrGlnCysThiLeuLeuCysProProArg 179  
Db 552371 GAACGCTTGGCCACCTTAATGCAAAATTAGGTGGCGAAGTGTATATCTCAACCTCAA 552312  
Qy 180 LeuCysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGly 199  
Db 552311 TTTTGTACAGATTAATGTCGATGATGCTTACACAGGTTTATTAACGTTAAACAAAGGT 552252  
RESULT 11  
US-09-643-990A-1/c  
; Sequence 1, Application US/09643990A  
; Patent No. 6528289  
; GENERAL INFORMATION:  
; APPLICANT: Robert D. Fleischmann  
; Mark D. Adams  
; Owen White  
; Hamilton O. Smith

```
J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS V6.22
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1
Alignment Scores:
Pred. No.: 7.8e-15 Length: 1830121
Score: 224.00 Matches: 67
Percent Similarity: 47.50% Conservative: 28
Best Local Similarity: 33.50% Mismatches: 83
Query Match: 18.06% Indels: 22
DB: 4 Gaps: 6
US-10-649-273-2_COPY_176_414 (1-239) x US-09-643-990A-1 (1-1830121)
QY 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21
Db 552791 TTAAGCGCGTGTGAGTGTAGAAAATAAGAGATGAGAGATCTATTGATGATGCT 552732
QY 22 ProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyCys 41
Db 552731 GCTGGCGAAGCCTTGTGATTAACACGCAAAATTAATTGACCTA-----GATTATCCA----- 552681
QY 42 SerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHis 61
Db 552680 -----GATGGCGCGCGACACTTCTGCTTTAGCGGAAAGGTACGCCCAATCGT 552633
QY 62 PheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGlyLeu 81
Db 552632 TTTCACATTTCCACGTCGCAATGACAGATCGTGCAGCGCTTGATTTTCTGCTTAA 552573
QY 82 GlnHisValThrAspLysIleIleMetLysGlnLysGlnGly-----IleGln 99
Db 552572 AAAACATTGCGCGCAATACAGTATATCAAGCAATTAAAGAGGCGCAACTGATAGAG 552513
QY 100 LysGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAla 119
```

```
Db 552512 CAA-----ACTAAGCAGATATTGCTTATGCTTTCCAAATGCGGTGG 552468
QY 120 CysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeu 139
Db 552467 GATACTCTTGCC-----ATTAATGTAAAG---CGTCATTTGAAA 552432
QY 140 ProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlaArg 159
Db 552431 GAAACAGCGCTATAAGCTTATGATGATTCGGGAGGGGTGAGCGCAATATAAACTCCCA 552372
QY 160 ArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPArg 179
Db 552371 GAAACGCTTGGCCGACCTTAATGCAAAATTTAGGTGGGAGGTGTTTATCTCAACCTCA 552312
QY 180 LeuCyThrAspAsnGlyLysIleMetIleAlaTPAsnGlyLysGlnArgLeuArgIaGly 199
Db 552311 TTTTGTACAGATTAATGTTGCGATGATGCTTACACAGGTTTTTTTACGTTTAAACAAAGGT 552252
RESULT 12
US-09-902-540-6612
Sequence 6612, Application US/09902540
Patent No. 6833447
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 6612
LENGTH: 996
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-6612
Alignment Scores:
Pred. No.: 8.47e-20 Length: 996
Score: 221.50 Matches: 70
Percent Similarity: 47.50% Conservative: 25
Best Local Similarity: 35.00% Mismatches: 74
Query Match: 17.86% Indels: 31
DB: 4 Gaps: 8
US-10-649-273-2_COPY_176_414 (1-239) x US-09-902-540-6612 (1-996)
QY 5 ValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAsp 24
Db 433 GTGCAAGCGCTACCGGCGAGTAACCGCTGTGGGCGAGCGACACACGCGCGCGAG 492
QY 25 MetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThrMet 44
Db 493 GCATATGACAAAGCCGCTGCATCTCGGCTG---CCGATCCG----- 534
QY 45 SerGlyLysValAlaIleGlnHisLeuAlaLysGlnGlyAsn-----ArgPhe 60
Db 535 ---GATGGCGAGCCCATGACACAGTTGGCGGCGAGGGAACCGGAGGCGCATCCGCTTC 591
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
Db 592 -----CCGCGCGCGCTGCGGCGGCAACAATTCCAGCGTCTCTTCGCGG 636
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlnLysGlnGlyIleGlnLys 100
Db 637 TTGAAG-----ACGGGCGTGCACACAGTGCAGAGGACGGCGTGCAGAG 684
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
```

```
Db 685 GGGAGGCGCTG-----GCGATTTCGCGGCTCTTCAGAGGCGGTGGCGAAC 735
Qy 121 HisLeuVallysaArgThrHisArgAlaIleLeuPheCysylsGlnArgAspLeuPro 140
Db 736 GTGCTGTGCAAGAG-----CTGTGGCGCGCGCGCGCGGTG----- 774
Qy 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160
Db 775 ---GGCCACAGACAGATTGTGTCTGTGGCGCGCGCTGCGCGCACTCGCGGTGGCGGCA 831
Qy 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db 832 CTGTGTCAAGCGCGACCGGAGCGGGGTTGAACATGTTCTGCCCCGGTGGCGCTG 891
Qy 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 892 TGCACGGACATATGGCGCATATTCGGTGGCGGGTATGAGCGGATCCGCGCGGCGCTG 951

RESULT 13
US-09-540-503/c
; Sequence 503, Application US/0902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10115849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 503
; LENGTH: 2582
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-503

Alignment Scores:
Pred. No.: 3,98e-19 Length: 2582
Score: 221.50 Matches: 70
Percent Similarity: 47.50% Conservative: 25
Best Local Similarity: 35.00% Mismatches: 74
Query Match: 17.86% Indels: 31
DB: Gaps: 8

US-10-649-273-2_COPY_176_414 (1-239) x US-09-902-540-503 (1-2582)
Qy 5 ValGlnIlyValSerAspPheLeuLeuGlyIlySerLeuAspIleAlaProGlyAsp 24
Db 1096 GTGAGAGCTTAACGGGACGTACCGGCTGTGTGGCAGACACCGGAGACCGCGCGAG 1037
Qy 25 MetLeuAspIlyValAlaArgArgLeuSerLeuIleIlyshIshProGlyCysSerThrMet 44
Db 1036 GCATATACAAAGACCGCTCGATCTCGGCGCTG---CCGTATCCG----- 995
Qy 45 SerGlyIlyValAlaIleGlnHisLeuAlaIlysgInGlyAsn-----ArgPhe 60
Db 994 ---GTGTGGCAGCGCATGACAGTTGGCGGAGGAGAACCCGAGGCCATCGGCTTC 938
Qy 61 HisPheAspIleIlyshProProLeuHisHisAlaIlyAsnCysAspPheSerPheThrGly 80
Db 937 -----CCGCGCGCGCTGCGCGGCGCAACTTCAGTCTCTTCCTCCGCGG 893
Qy 81 LeuGlnHisValThrAspIlyIleIleMetIlyshGlyIlysgInGlyIleGlnIly 100
Db 892 TTAAAG-----ACGCGGCTGTGCAACCGTGCAGAACCGCGGCTGTGCGGCA 845
Qy 101 GlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120
Db 844 GGGCAGGCGCTG-----GCGATTTCGCGGCTCTTCAGAGGCGCGTGGCGAAC 794
```

```
Qy 121 HisLeuVallysaArgThrHisArgAlaIleLeuPheCysylsGlnArgAspLeuPro 140
Db 793 GTGCTGTGCAAGAG-----CTGTGGCGCGCGCGCGCGGTG----- 755
Qy 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160
Db 754 ---GGCCACAGACAGATTGTGTCTGTGGCGCGCGCTGCGCGCAACTCGCGGTGGCGGCA 698
Qy 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
Db 697 CTGTGTCAAGCGCGACCGGAGCGGGGTTGAACATGTTCTGCCCCGGTGGCGCTG 638
Qy 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 637 TGCACGGACATATGGCGCATATTCGGTGGCGGGTATGAGCGGATCCGCGCGGCGCTG 578

RESULT 14
US-09-543-681A-2341
; Sequence 2341, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2341
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2341

Alignment Scores:
Pred. No.: 1,86e-18 Length: 1074
Score: 212.00 Matches: 66
Percent Similarity: 44.95% Conservative: 32
Best Local Similarity: 30.28% Mismatches: 90
Query Match: 17.10% Indels: 30
DB: Gaps: 7

US-10-649-273-2_COPY_176_414 (1-239) x US-09-543-681A-2341 (1-1074)
Qy 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyIlySerLeuAspIleAla 21
Db 475 TTAAATTAGTGAACAGGATTTGGGGAATATACCTGTTAGTGAATGATGATGCT 534
Qy 22 ProGlyAspMetLeuAspIlyValAlaArgArgLeuSerLeuIleIlyshIshProGlyCys 41
Db 535 GCTGTGAAGCAATTGATGAATAACAGCAAGCATTTGGGGCTT---GATTATCCG----- 585
Qy 42 SerThrMetSerGlyIlyValAlaIleGlnHisLeuAlaIlysgInGlyAsnArgPheHis 61
Db 586 -----GCGCGCGCTGTTTATCAAAAATGCGCAACAAAGTGTAGAGGACGCT 633
Qy 62 PheAspIleIlyshProProLeuHisHisAlaIlyAsnCysAspPheSerPheThrGlyLeu 81
Db 634 TTTGTTTTCTCGTCCCATGACAGACAGACCGGACGTGACCTTTAGTTCTCTGTGTTA 693
Qy 82 GlnHisValThrAspIlyIleIleMetIlyshGlyIlysgInGlyIleGlnIly 101
Db 694 AAACCTTGGCGGCTAATGATATTCGTAACAAAGATGATTCAGAG----- 738
Qy 102 GlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHis 121
Db 739 -----CAAACTCGAGCATATTCGCGCTGCTTTGAAGATCCGTAAGATAGTACT 789
Qy 122 LeuVallysaArgThrHisArgAlaIleLeuPheCysylsGlnArgAspLeuProGln 141
```

Db 790 TTGGCAATAAATGTCGACGA-----TTAGAGCA 822  
Qy 142 AsnAn---AlaValLeuValAlaSerGlyValAlaSerAspPheTyrIleArg 160  
Db 823 ACAGCCTTTAAACCTTAGTATGATGCTGGGGCGTAAGTCTAACCGTAACCTTAGCGCC 882  
Qy 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
Db 883 AAATGGGATGATGAATGAACAACCTCGAGGGGAGAGTGTATTATGCTGCGCCCTGAGTTA 942  
Qy 181 CysThrAspAsnGlyIleMetIleAlaTTPraSngIyIleGluArgLeuArgAlaGlyLeu 200  
Db 943 TGTACCGATTAATGATGTCATATGCTTGGCGGAGATGATCCGTTTAAAGTGTACC 1002  
Qy 201 ---GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeu 217  
Db 1003 GAGGAGCCTTTA-----GAGGTGACAGTGAACACAGCTTGGCCTTTA 1044

RESULT 15  
US-09-489-039A-2050  
; Sequence 2050, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 2050  
; LENGTH: 1032  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-2050

Alignment Scores:  
Pred. No.: 1.55e-17 Length: 1032  
Score: 205.00 Matches: 66  
Percent Similarity: 43.58% Conservative: 29  
Best Local Similarity: 30.28% Mismatches: 93  
Query Match: 16.53% Indels: 30  
DB: 4 Gaps: 7

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-09-489-039A-2050 (1-1032)

Qy 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21  
Db 442 TTGATTACCGTACACCGGTAATGTCATATGTAACCTGCGGAGTCGATTACGATGCG 501  
Qy 22 ProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCys 41  
Db 502 GCGGGCGAAGCCTTGTATGAAGACGCGAAGCTGTGGACTG--GATTATCCC----- 552  
Qy 42 SerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHis 61  
Db 553 -----GGCGGGCGGATGCTGTGAAATGCGCTCCAGGGCGACCGAAGGCCGC 600  
Qy 62 PheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeu 81  
Db 601 TTGTCTTCCCGCGCGCATGACCGACCTCGGGGCTGAGCTTCAGCTTCGCGCTG 660  
Qy 82 GlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLysGly 101  
Db 661 AAGACCTTGGCGGCAACACATTCCGACGCAACGGGAGCATGAG----- 705  
Qy 102 GlnIleuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHis 121  
Db 706 -----CAAACCGCGCGCATGACCGCGGCGCTTGGAGATGCGGTCGATACG 756  
Qy 122 LeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGln 141

Db 757 CTGATGATTAAATGTCGCGCGCG-----CTGAGACAA 789  
Qy 142 AsnAn---AlaValLeuValAlaSerGlyValAlaSerAspPheTyrIleArg 160  
Db 790 ACCGCTTTAAGCGTGTGATGCGGGAGAGCGTAAGCCCAACCTTACCTCGCGCG 849  
Qy 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
Db 850 AAGCTGGCGGAGATGATGCAAAACGCGCGGAGTGTATTACCGCCCTGAGTTG 909  
Qy 181 CysThrAspAsnGlyIleMetIleAlaTTPraSngIyIleGluArgLeuArgAlaGlyLeu 200  
Db 910 TGTACTGACAAACGCGCGATGATGCGCTAACGCGGCGATGATGCTGCAAAACCGCGCG 969  
Qy 201 GlyIleLeuHisAspIleGlu---GlyIleArgTyrGluProLysCysProLeu 217  
Db 970 -----AAAGCGAGCTCGGCGTGAACGCTTGGCCGCGCTGGCGCTG 1011

Search completed: November 11, 2005, 03:17:05  
Job time : 1010.64 secs

Gencore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 10, 2005, 16:59:00 ; Search time 793.636 Seconds  
(without alignments)  
2490.420 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 LIALVGVSDFLILGKSLDI.....DISKVGESAIKVPQLMREI 239

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xid  
-O/cgn2\_1/USFTO\_spool\_p/US10649273/runat\_02112005\_091339\_15609/app\_query.fasta\_1.1429  
-DB=Published Applications\_NA -QFMT=fastdp -SUFFIX=p2n.rmpb -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.csl -LIST=45 -DOCCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10649273 @CGN 1.1 1034 @runat 02112005\_091339\_15609  
-NCPU=6 -ICPU=3 -NO MMAP -LARGESOURC -NEG SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FCAPOP=6 -FCAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq.\*  
20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq.\*  
21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq.\*  
22: /cgn2\_6/ptodata/1/pubpna/US10J\_PUBCOMB.seq.\*  
23: /cgn2\_6/ptodata/1/pubpna/US10K\_NEW\_PUB.seq.\*  
24: /cgn2\_6/ptodata/1/pubpna/US10L\_NEW\_PUB.seq.\*  
25: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq.\*  
26: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*  
27: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
28: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1240	100.0	1416	18	US-10-120-988-177
2	1240	100.0	1526	15	US-10-067-443-23
3	1240	100.0	1526	19	US-10-649-273-23
4	1240	100.0	1526	19	US-10-651-722-23
5	1240	100.0	2197	15	US-10-067-443-1
6	1240	100.0	2197	19	US-10-649-273-1
7	1240	100.0	2197	19	US-10-651-722-1
8	1240	100.0	2572	22	US-10-480-988-16
9	1213.5	97.9	1387	15	US-10-067-443-21
10	1213.5	97.9	1387	19	US-10-649-273-21
11	1213.5	97.9	1387	19	US-10-651-722-21
12	1213.5	97.9	1387	22	US-10-887-553A-1047
13	1203	97.0	1245	15	US-10-012-140-6
14	1203	97.0	1820	15	US-10-012-140-6
15	1059	85.4	2208	18	US-10-094-749-400
16	1059	85.4	2890	21	US-10-723-860-7447
17	725	58.5	14364	15	US-10-067-443-20
18	725	58.5	14364	19	US-10-649-273-20
19	725	58.5	14364	19	US-10-651-722-20
20	599	48.3	2734	24	US-10-450-763-20426
21	468	37.7	371	20	US-10-430-201-3118
22	468	37.7	371	20	US-10-430-201-3119
23	337	27.2	1917	19	US-10-424-599-66417
24	337.5	26.8	1601	26	US-11-097-143-34190
25	332.5	26.8	3656	26	US-11-097-143-34189
26	270	21.8	1628	20	US-10-437-963-11249
27	257	20.7	1146	18	US-10-282-122A-11249
28	248.5	20.0	1000	20	US-10-343-561-50
29	248.5	20.0	1044	18	US-10-282-122A-26972
30	248.5	20.0	94750	19	US-10-672-787-18
31	247	19.9	936	18	US-10-282-122A-8315
32	241	19.4	1032	18	US-10-282-122A-31043
33	239	19.3	756	15	US-10-081-051-8
34	239	19.3	4360	15	US-10-081-051-2
35	234	18.9	300	24	US-10-779-543-7021
36	231	18.6	1026	9	US-09-815-242-7701
37	231	18.6	1026	18	US-10-282-122A-30016
38	231	18.6	1026	24	US-10-958-216-49
39	229	18.5	1026	24	US-10-958-216-51
40	224	18.1	1029	9	US-09-815-242-6946
41	224	18.1	1029	18	US-10-282-122A-22020
42	224	18.1	1830121	18	US-10-329-670-1
43	224	18.1	1830121	21	US-10-158-865-1
44	224	18.1	1830121	24	US-10-981-687-1
45	217	17.5	9967	24	US-10-795-159-529

ALIGNMENTS

RESULT 1  
US-10-120-988-177  
; Sequence 177, Application US/10120988  
; Publication No. US20030219745A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Liu, Chenhua  
; APPLICANT: Ren, Feiyun  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Dmanac, Radoje T.  
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and  
; FILE REFERENCE: 802CON

```

; CURRENT APPLICATION NUMBER: US/10/120,988
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 09/774,528
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pc_FL_genes Version 2.0
; SEQ ID NO 177
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205) ..(1305)
US-10-120-988-177

Alignment Scores:
Pred. No.: 2,38e-151 Length: 1416
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 18 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-120-988-177 (1-1416)
QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 586 CTGTTGGCATTACTTCAAGAGATTTCAGATTTCTGCTTGGAAAAGCTTTGGACATA 645
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 646 GCACCGAGTGACATGCTTGAACAAGTGCCAAAGACCTTTCTTAATAAATGATCCAGAG 705
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGInGlyAsnArgPhe 60
DB 706 TGGTCCACCATGAGTGGTGGAAAGCCATGAAACATTGGCCAAACAAAGAAATGATTT 765
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 766 CATTTGACATCAAACTCCCTTGCACTGCTTAATAATTTGATTTTCTTTACTGGA 825
QY 81 LeuGInHisValThrAspLysIleIleMetLysGlyLysGluLysGluGlyIleGluLys 100
DB 826 CTTCAACACGTTACTGATAAATATATATGAAAAAGAGAAAGGATTTGAGAG 885
QY 101 GlyGInIleLeuSerSerAlaAlaAspIleAlaIleThrValGInHisThrMetAlaCys 120
DB 886 GGGCAATCTGCTTTCAGACAGACAGATGCTCCACAGTACACACATGCAATGCAATGT 945
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
DB 946 CATCTGTGAAAAGAACATCGGGCTATTTCTGTTGTATGACAGAGACTTTTACT 1005
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
DB 1006 CAAAATATATGCAATGATGCTGATCTGGTGTCTCCAGTAACTTCTATATCCGAGA 1065
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
DB 1066 GCTCTGAAATTTTAAACAAACGCAACAGTGCATTTGTGTCTCCCTCCACAGACTA 1125
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1126 TGCACTGATATATGCAATATGATGATGATGATGATGATGATGATGATGATGATGATG 1185
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
DB 1186 GGCATTTTATATGACATATGAAAGGATCCGCTATGATACCAAAATGTCCTCTTGAGTAC 1245
QY 221 HisSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB 1246 ATATCAAAAAGAGTGGAGAGCTTCATATAAGTACACAAATTAATAATGAGATA 1302
```

```

RESULT 2
US-10-067-443-23
; Sequence 23, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-23

Alignment Scores:
Pred. No.: 2.65e-151 Length: 1526
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-23 (1-1526)
QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 85 CTGTTGGCATTACTTCAAGAGATTTCAGATTTCTGCTTGGAAAAGCTTTGGACATA 144
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 145 GCACCGAGTGACATGCTTGAACAAGTGCCAAAGACCTTTCTTAATAAATGATCCAGAG 204
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGInGlyAsnArgPhe 60
DB 205 TGGTCCACCATGAGTGGTGGAAAGCCATGAAACATTGGCCAAACAAAGAAATGATTT 264
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 265 CATTTGACATCAAACTCCCTTGCACTGCTTAATAATTTGATTTTCTTTACTGGA 324
QY 81 LeuGInHisValThrAspLysIleIleMetLysGlyLysGluLysGluGlyIleGluLys 100
DB 325 CTTCAACACGTTACTGATAAATATATATGAAAAAGAGAAAGGATTTGAGAG 384
QY 101 GlyGInIleLeuSerSerAlaAlaAspIleAlaIleThrValGInHisThrMetAlaCys 120
DB 385 GGGCAATCTGCTTTCAGACAGACAGATGCTCCACAGTACACACATGCAATGCAATGT 444
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
DB 445 CATCTGTGAAAAGAACATCGGGCTATTTCTGTTGTATGACAGAGACTTTTACT 504
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
DB 505 CAAAATATATGCAATGATGCTGATCTGGTGTCTCCAGTAACTTCTATATCCGAGA 564
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
DB 565 GCTCTGAAATTTTAAACAAACGCAACAGTGCATTTGTGTCTCCCTCCACAGACTA 624
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 625 TGCACTGATATATGCAATATGATGATGATGATGATGATGATGATGATGATGATGATG 684
```

```
OY      201 GYIleuHISAspIleGluGlyIleArgTYrGluProLysCysProLeuGlyValAsp 220
DB      685 GGCATTTCATGACATAGAGGATCCGCTATGAAACCAAAATGCTCTTGAGATGAC 744
OY      221 ILeSerLySGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB      745 ATATCAAAAGAAAGTTGAGAAAGCTTCATTAAGTACCAACATTAATAATGGAGATA 801

RESULT 3
US-10-649-273-23
; Sequence 23, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CMT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-649-273-23

Alignment Scores:
Pred. No.:      2,65e-151      Length:      1526
Score:          1240.00      Matches:      239
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:            19          Gaps:      0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-649-273-23 (1-1526)
OY      1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIle 20
DB      85 CTGTGGCATTAAGTTCAGAGGATTTTCAGATTTCTGCTTCGAAAGTCTTGGACATA 144
OY      21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB      145 GCACCAAGGTGACATGCTTGCAAGGTGGCAAGAACCTTCTTAATAAATCATCCAGAG 204
OY      41 CysSerThrMetSerGlyLysValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB      205 TGCCTCCACCAAGAGTGTGGGAAAGCCATGAAACATTGGCCAAACAGAAATAGATTT 264
OY      61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB      265 CATTTCGACATCAAACTCCCTTCGATCATGCTTAATAATTTGATTTTCTTTTACTGGA 324
OY      81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100
DB      325 CTTCAACACGTTACTGATTAATAATATGAAAAAGAAAAAGAGGAAGGATTTGAGAG 384
OY      101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB      385 GGGCAAAATCTGCTTCACAGACAGACATTGCTGCCACAGTACGACACACATGCGCATGT 444
OY      121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnAspAspLeuLeuPro 140
DB      445 CATCTTGTAAGAAAGACATACGGGCTATTCTGTTTGTAAGCAGAGAGACTGTATCACT 504
OY      141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTYrIleArgArg 160
DB      505 CAAATAATATGCACTACTGTTGATCTGTGTGTCTGCCAAGTACTTCTAATATCCGACGA 564
```

```
OY      161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuCysProProArgLeu 180
DB      565 GCTCTGCAAAATTTTAAACAAACGACACAGTGCATCTTGTGTGCTCTCCACAGATA 624
OY      181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB      625 TGCATGTAAATGAGCATTAATGATTCATGGAATGGATGTAAGAAAGTACGTGCTGCTTG 684
OY      201 GlyIleLeuHisAspIleGluGlyIleArgTYrGluProLysCysProLeuGlyValAsp 220
DB      685 GGCATTTCATGACATAGAGGATCCGCTATGAAACCAAAATGCTCTTGAGATGAC 744
OY      221 ILeSerLySGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB      745 ATATCAAAAGAAAGTTGAGAAAGCTTCATTAAGTACCAACATTAATAATGGAGATA 801

RESULT 4
US-10-651-722-23
; Sequence 23, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-651-722-23

Alignment Scores:
Pred. No.:      2,65e-151      Length:      1526
Score:          1240.00      Matches:      239
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:            19          Gaps:      0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-651-722-23 (1-1526)
OY      1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIle 20
DB      85 CTGTGGCATTAAGTTCAGAGGATTTTCAGATTTCTGCTTCGAAAGTCTTGGACATA 144
OY      21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB      145 GCACCAAGGTGACATGCTTGCAAGGTGGCAAGAACCTTCTTAATAAATCATCCAGAG 204
OY      41 CysSerThrMetSerGlyLysValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB      205 TGCCTCCACCAAGAGTGTGGGAAAGCCATGAAACATTGGCCAAACAGAAATAGATTT 264
OY      61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB      265 CATTTCGACATCAAACTCCCTTCGATCATGCTTAATAATTTGATTTTCTTTTACTGGA 324
OY      81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100
DB      325 CTTCAACACGTTACTGATTAATAATATGAAAAAGAAAAAGAGGAAGGATTAATGAGAG 384
OY      101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
```



```
Db 385 GGGCAAAATCTGTTTCAGCAGACAGACATTGCTCCACAGTACACACAAATGGCATGT 444
Qy 121 HisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuPro 140
Db 445 CATCTTGAAAAGAACACATCGGGCTATCTGTTTGTAAAGCAGAGACTTGTACTT 504
Qy 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
Db 505 CAATAATATGCAGACTGCTGTTGCACTGGTGCTGCCAAGTAACTTCTATATCCGACGA 564
Qy 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
Db 565 GCTCTGGAATTTTAAACAAACGCAACAGTGCACCTTGTGTGCTCCCTCCACAGACTA 624
Qy 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 625 TGCACCTGATATGGCATTTATGATTCATGATGATGATGATGATGATGATGATGATGATG 684
Qy 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProIlyCysProLeuGlyValAsp 220
Db 685 GGCATTTTACATGACATAGAAAGCATCCGCTATGACCAAAATGTCCCTTGGAGTAAAC 744
Qy 221 ILeSerIysGluValGlyGluAlaSerIleIysValProGlnLeuIlyMetGluIle 239
Db 745 ATATCAAAAAGAGTTGAGAGAGCTTCATATAAAGTACCAATTAATAATGAGATA 801

RESULT 5
US-10-067-443-1
; Sequence 1, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED I
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-067-443-1

Alignment Scores:
Pred. No.: 4,52e-151 Length: 2197
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
15

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-1 (1-2197)
Qy 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyIysSerLeuAspIle 20
Db 756 CTGTGGCATATGATTCAGAGAGATTTTCAGATTTTCGCTTCTTGGAAGATCTTTGACATA 815
Qy 21 AlaProGlyAspMetLeuAspIysValAlaArgArgLeuSerLeuIleIlyHisProGlu 40
Db 816 GCACCAAGGTGACATGCTTGAAGGTGGCAAGAACATCTTTTAATAAATCATCCAGAG 875
Qy 41 CysSerThrMetSerGlyGlyIysAlaIleGluHisLeuAlaIysGlnGlyAsnArgPhe 60
Db 876 TGCCTCCACACAGATGATGGAGAAAGCCATAGACATTTGGCCAAACAGAAATGATTT 935
```

```
Qy 61 HisPheAspIleIysProProLeuHisHisAlaIysAsnCysAspPheSerThrGly 80
Db 936 CATTTGACATCAAACTCCCTTGATCATGTATAAAATGGATTTTCTTTTACTGGA 995
Qy 81 LeuGlnHisValIleThrAspIysGlyIleIleMetIysIysGluIysGluGlyIleGluIys 100
Db 996 CTTCAACCGTTACTGATTAATAATATATGAAAAAGAAAAAGAGAGATTTTAAAG 1055
Qy 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120
Db 1056 GGGCAAAATCTGTTCTTCAGCAGACAGATTGCTGCCACAGTACAGACACAAATGGCATGT 1115
Qy 121 HisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuPro 140
Db 1116 CATCTTGAAAAGAACACATCGGGCTATCTGTTTGTAAAGCAGAGACTTGTACTT 1175
Qy 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160
Db 1176 CAATAATATGACAGTACTGTTGCACTGGTGCTGCGCAAGTAACTTCTATATCCGACGA 1235
Qy 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
Db 1236 GCTCTGGAATTTTAAACAAACGCAACAGTGCACCTTGTGTGCTCCCTCCACAGACTA 1295
Qy 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 1296 TGCACCTGATATGGCATTTATGATTCATGATGATGATGATGATGATGATGATGATGATG 1355
Qy 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProIlyCysProLeuGlyValAsp 220
Db 1356 GGCATTTTACATGACATAGAAAGCATCCGCTATGACCAAAATGTCCCTTGGAGTAAAC 1415
Qy 221 ILeSerIysGluValGlyGluAlaSerIleIysValProGlnLeuIlyMetGluIle 239
Db 1416 ATATCAAAAAGAGTTGAGAGAGCTTCATATAAAGTACCAATTAATAATGAGATA 1472

RESULT 6
US-10-649-273-1
; Sequence 1, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-649-273-1

Alignment Scores:
Pred. No.: 4,52e-151 Length: 2197
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
19

US-10-649-273-2_COPY_176_414 (1-239) x US-10-649-273-1 (1-2197)
```

```
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAsp11e 20
Db 756 CTGTTGGCATTAGTTCAGAGAGCTTCAGATTTTCCTTCCTTGGAAGCTTTGGACATA 815
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeu11elysH1sProGlu 40
Db 816 GCACCAAGGTGACATCTTGACAAAGGTGGCAAGAAAGCTTTCTTAATTAACATCCAGAG 875
QY 41 CysSerThrMetSerGlyGlyLysAla11eGluH1sLeuAlaLysGlnGlyAsnArgPhe 60
Db 876 TGCTCCACCATGAGTGGTGGGAAGCCATAGAACTTTGGCCAAACAGAAATTAATTT 935
QY 61 HisPheAsp11elysProProLeuH1sHisAlaLysAsnCysAspPheSerPheThrGly 80
Db 936 CATTTTGACATCAACCTCCCTTGATCATGTCTAAATAATTTGATTTTCTTTTACTGGA 995
QY 81 LeuGlnH1sValThrAspLys11e11eMetLysLysGlnLysGlnGlyLys11eGluLys 100
Db 996 CTTCACACGTTACTGATTAATAATTAATGAAGAAAGAAAGAAAGATTAATGAGAG 1055
QY 101 GlyGln11eLeuSerSerAla11aAsp11eAla11aThrValGlnH1sThrMetAlaCys 120
Db 1056 GGGCAATCTCTGCTTCACGACACACATTCCTGCCACGTCACACACATGCGATGT 1115
QY 121 HisLeuValLysArgThrHisArgAla11eLeuPheCysLysGlnArgAspLeuPro 140
Db 1116 CATCTTGGAAGAAAGACATCGGGCTATTCTGTTTGTGACAGAGACATGTTACTGT 1175
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAla11aSerAsnPheYr11eArgArg 160
Db 1176 CAATAATATGACAGTACTGGTGTGATCTGGTGTGCGCAAGTAACTTCTATATCCGACA 1235
QY 161 AlaLeuGln11eLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
Db 1236 GCTCTGGAATTTTAAACAGCAACAGTCACCTTGTTGTGTCTCTCCACAGACTA 1295
QY 181 CysThrAspAsnGly11eMet11eAla11aThrAsnGly11eGlnArgLeuArgAlaGlyLeu 200
Db 1296 TGCACTGATTAATGCAATATGATTCATGGAATGATTAATGAAGACTACTGCTGCTTG 1355
QY 201 Gly11eLeuH1sAsp11eGlnGly11eArgYrGlnProLysCysProLeuGlyValAsp 220
Db 1356 GGCATTTTACATGACATGAGAGGCAATCCGCTATGAACCAAAATGTCTCTTGAGATGAGC 1415
QY 221 11eSerLysGlnValGlyGlnAlaSer11eLysValProGlnLeuLysMetGlu11e 239
Db 1416 ATATCAAAAGAGTTGAGAGCTTCATTAAGTACCACAATTAATAATGAGATA 1472

RESULT 7
US-10-651-722-1
; Sequence 1, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
```

```
; LOCATION: (231) .. (1472)
US-10-651-722-1
Alignment Scores:
Pred. No.: 4,52e-151 Length: 2197
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 19 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-651-722-1 (1-2197)
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAsp11e 20
Db 756 CTGTTGGCATTAGTTCAGAGAGCTTCAGATTTTCCTTCCTTGGAAGCTTTGGACATA 815
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeu11elysH1sProGlu 40
Db 816 GCACCAAGGTGACATCTTGACAAAGGTGGCAAGAAAGCTTTCTTAATTAACATCCAGAG 875
QY 41 CysSerThrMetSerGlyGlyLysAla11eGluH1sLeuAlaLysGlnGlyAsnArgPhe 60
Db 876 TGCTCCACCATGAGTGGTGGGAAGCCATAGAACTTTGGCCAAACAGAAATTAATTT 935
QY 61 HisPheAsp11elysProProLeuH1sHisAlaLysAsnCysAspPheSerPheThrGly 80
Db 936 CATTTTGACATCAACCTCCCTTGATCATGTCTAAATAATTTGATTTTCTTTTACTGGA 995
QY 81 LeuGlnH1sValThrAspLys11e11eMetLysLysGlnLysGlnGlyLys11eGluLys 100
Db 996 CTTCACACGTTACTGATTAATAATTAATGAAGAAAGAAAGAAAGATTAATGAGAG 1055
QY 101 GlyGln11eLeuSerSerAla11aAsp11eAla11aThrValGlnH1sThrMetAlaCys 120
Db 1056 GGGCAATCTCTGCTTCACGACACACATTCCTGCCACGTCACACACATGCGATGT 1115
QY 121 HisLeuValLysArgThrHisArgAla11eLeuPheCysLysGlnArgAspLeuPro 140
Db 1116 CATCTTGGAAGAAAGACATCGGGCTATTCTGTTTGTGACAGAGACATGTTACTGT 1175
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAla11aSerAsnPheYr11eArgArg 160
Db 1176 CAATAATATGACAGTACTGGTGTGATCTGGTGTGCGCAAGTAACTTCTATATCCGACA 1235
QY 161 AlaLeuGln11eLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
Db 1236 GCTCTGGAATTTTAAACAGCAACAGTCACCTTGTTGTGTCTCTCCACAGACTA 1295
QY 181 CysThrAspAsnGly11eMet11eAla11aThrAsnGly11eGlnArgLeuArgAlaGlyLeu 200
Db 1296 TGCACTGATTAATGCAATATGATTCATGGAATGATTAATGAAGACTACTGCTGCTTG 1355
QY 201 Gly11eLeuH1sAsp11eGlnGly11eArgYrGlnProLysCysProLeuGlyValAsp 220
Db 1356 GGCATTTTACATGACATGAGAGGCAATCCGCTATGAACCAAAATGTCTCTTGAGATGAGC 1415
QY 221 11eSerLysGlnValGlyGlnAlaSer11eLysValProGlnLeuLysMetGlu11e 239
Db 1416 ATATCAAAAGAGTTGAGAGCTTCATTAAGTACCACAATTAATAATGAGATA 1472

RESULT 8
US-10-480-988-36
; Sequence 36, Application US/10480988
; Publication No. US20050069877A1
; GENERAL INFORMATION:
; APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;
; APPLICANT: SWARNAKAR, Anita; HARPALIA, Apri1 J.A.;
; APPLICANT: TRAN, Bao; DUGGAN, Brendan M.;
; APPLICANT: WARREN, Bridget A.; ISON, Craig H.;
; APPLICANT: HONCHELU, Cynthia D.; NGUYEN, Daniel B.;
; APPLICANT: LU, Dzung Aina M.; LEB, Ernestine A.;
; APPLICANT: YUB, Henry; FORSYTHE, Ian J.;
```

```
APPLICANT: BARROSO, Ines; RAMKUMAR, Jayalaxmi;
APPLICANT: GRIFFIN, Jennifer A.; Li, Joana X.;
APPLICANT: YANG, Junming; THANGAVELU, Kavitha;
APPLICANT: GIETZEN, Kimberly J.; DING, Li;
APPLICANT: BAUGHN, Marian R.; BOROMSKY, Mark L.;
APPLICANT: YAO, Monique G.; CHAMLA, Nandinder K.;
APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;
APPLICANT: LEE, Sally; BECHTA, Shanya D.;
APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;
APPLICANT: ELLIOTT, Vicki S.; LEO, Wen;
APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;
APPLICANT: LU, Yan; ZEBARJADIAN, Yeganeh
TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
FILE REFERENCE: PF-1040 USN
CURRENT APPLICATION NUMBER: US/10/480,988
CURRENT FILING DATE: 2003-12-16
PRIOR APPLICATION NUMBER: PCT/US02/19360
PRIOR FILING DATE: 2002-06-18
PRIOR APPLICATION NUMBER: US 60/300,508
PRIOR FILING DATE: 2001-06-22
PRIOR APPLICATION NUMBER: US 60/303,445
PRIOR FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: US 60/305,405
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/311,442
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 60/314,821
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/315,992
PRIOR FILING DATE: 2001-08-29
PRIOR APPLICATION NUMBER: US 60/378,205
PRIOR FILING DATE: 2002-05-03
NUMBER OF SEQ ID NOS: 56
SOFTWARE: PERL Program
SEQ ID NO 36
LENGTH: 2572
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 7632424CB1
US-10-480-988-36

Alignment Scores:
Pred. No.: 5.7e-151 Length: 2572
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-480-988-36 (1-2572)
QY 1 LeuLeuAlaLeuValGInGInValISerAspPheLeuLeuGInGInValysSerLeuAspIle 20
DB 669 CTTGGCATTGAGTTCAAGAGTTTCAGATTTCCTCTTGGAAAGTTTGGACATA 728
QY 21 AlaProGlyAspMetLeuAspIysValAlaArgArgLeuSerLeuIleYsHISProGlu 40
DB 729 GCACCAAGTGCATGTTGACAAAGTGGCAAGAGACTTTCTTAATAAACAATCCAGAG 788
QY 41 CysSerThrMetSerGlyGlyValAlaIleGInHISLeuAlaIysGInGInValysnArgPhe 60
DB 789 TGCTCCACCATGATGTGGGAAAGCCATAGAACTTTGGCCAAACAGAAATAGATT 848
QY 61 HisPheAspIleYsProProLeuHISHisAlaYsAsnCysAspPheSerPheThrGly 80
DB 849 CATTTCACATCAAACTCCCTTGCAATGCTAAATAATTTGATTTTCTTTTACCTGGA 908
QY 81 LeuGInHISValThrAspIysIleIleMetIysGInGInGInGInGInGInGInGInGIn 100
DB 909 CTTCAACACGTTTACTATAATATATATATATATATATATATATATATATATATATATAT 966
```

```
QY 101 GInGInIleLeuSerSerAlaAlaAspIleAlaIleValGInHISThrMetAlaCys 120
DB 969 GGGCAAAATCTGTCCTTACAGACGACATTTGGCCACAGTACAGACCAATGGCATGT 1028
QY 121 HisLeuValIysArgThrHisArgAlaIleLeuPheCysIysGInArgAspLeuPro 140
DB 1029 CATTCTGTATAAAGAACACATCGGCTATCTGTTTGTGAAGAGACTTGTTACCT 1088
QY 141 GInAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheYrIleArgArg 160
DB 1089 CAATAATATGCACTACTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1148
QY 161 AlaLeuGInIleLeuThrAsnAlaThrGInCysThrLeuLeuCysProProArgLeu 180
DB 1149 GCTCTGGAATTTTAAACAAGCAACAGTGACCTTGTGTGCTCTCTCCACAGCTA 1208
QY 181 CysThrAspAsnGlyIleMetIleAlaTPAsnGlyIleGInArgLeuAlaGlyLeu 200
DB 1209 TGCACCTGATTAATGCAATTGATTCATGGAATGTAATGAAAGACTACGTCGCTTG 1268
QY 201 GInIleLeuHISAspIleGInGInIleArgGlyIleArgGlyProIysCysProLeuGInValAsp 220
DB 1269 GGCATTTTACATGACATPAGAAGCATCCGCTATGAACCAAAATGCTCTTGGAGTAC 1328
QY 221 IleSerIysGInValGInGInAlaSerIleYsValProGInLeuYsMetGInIle 239
DB 1329 ATATCAAAAGAACTTGGAGAGCTTCATTAATGTAACAATTAATAATGAGAGATA 1385

RESULT 9
US-10-067-443-21
; Sequence 21, Application US/10067443
; Publication No. US2003082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-21

Alignment Scores:
Pred. No.: 6.82e-148 Length: 1387
Score: 1213.50 Matches: 238
Percent Similarity: 90.15% Conservative: 0
Best Local Similarity: 90.15% Mismatches: 1
Query Match: 97.86% Indels: 25
DB: 15 Gaps: 1

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-21 (1-1387)
QY 1 LeuLeuAlaLeuValGInGInValISerAspPheLeuLeuGInGInValysSerLeuAspIle 20
DB 549 CTTGGCATTGAGTTCAAGAGTTTCAGATTTCCTCTTGGAAAGTTTGGACATA 608
QY 21 AlaProGlyAspMetLeuAspIysValAlaArgArgLeuSerLeuIleYsHISProGlu 40
DB 609 GCACCAAGTGCATGTTGACAAAGTGGCAAGAGACTTTCTTAATAAACAATCCAGAG 668
QY 41 CysSerThrMetSerGlyGlyValAlaIleGInHISLeuAlaIysGInGInGInGInGInGIn 60
DB 669 TGCTCCACCATGATGTGGGAAAGCCATAGACATTTGGCCAAACAGAAATAGATT 728
```

QY	61	HisPheAsp11eLysProCProLeuH1sH1sAlaLysAsnCysAspPheSerPheTrgLY	80
Db	729	CATTTGACATCAACACCTCCCTGCATCATCGCAAAAATTTGCAATTTTCTTTTACATGGA	788
QY	81	LeuGlnH1sValThrAspLys1le1leMeLysLysGluLysGluGlu1yle-----	98
Db	789	CTTCAACACGTTACTGATTAATAATTAAGAAAAAGAAAAAGAGAGATATATTCTTA	848
QY	98	-----	98
Db	849	ATTAGTAAAGTTGAAACAGATTAATATTTCTGATTTGCTTAAAAATAGCTGCTCATTTTC	908
QY	99	-----GluLysGlyGln1leLeuSerSerAla1Asp11eAla1a1aThrValGln	115
Db	909	TGCAGGTATGAGAAAGGGCAATTCCTGCTTTACGACGACGACATTCCTCCACAGTACAG	968
QY	116	HisThrMetAla1CysHisLeuVal1LysAArgThrHisArgAla11leuPheCysLysGln	135
Db	969	CACACAAATGGCATGTCATCTTGTAAGAAAGAACATCGGGCATTTCTGTTTGTGAAGCAG	1022
QY	136	ArgAspLeuLeuProGlnAsnAsnAlaVal1leuValAlaSerGlyGlyVal1AsnAsn	155
Db	1029	AGAGACTTGTAACCTCAAAATATATCAGTACGTGCTGTCATCTGCTGCTCCCAAGTAAAC	1088
QY	156	PheTrp1LeArgArgAla1LeuGln1leLeuThrAsnAla1a1aThrGlnCysThrLeuLeuCys	175
Db	1089	TTCTATATCCGACAGACTCTGCAAAATTTTAAACAAGCAACACAGACACTTGTGTGTGT	1144
QY	176	ProProProArgLeuCysThrAspAsnGly1leMe1leAla1a1aTrpAsnGly1leGluArg	195
Db	1149	CCTCCTCCACAGACTATGACATGATATGCGCATTTATGATTTGATGATGATGATGATGAAGA	1200
QY	196	LeuArgAlaGlyLeuGly1yleuH1sAsp11eGluGly1yleArgTrpGluProLysCys	215
Db	1209	CTACCTGTGGCTTGCGCATTTTATCATGACATAGAACGATCCGCTATGAACCAAAATGT	1266
QY	216	ProLeuGlyValAsp1leSerLysGlyVal1GlyVal1aSer1leLysVal1ProGlnLeu	235
Db	1269	CCTCTGGAGTAGACATATCAAAAGAAAGTGGAGAAAGCTTCATTAAGTACCACATTA	1322
QY	236	LysMetGlu1le 239	
Db	1329	AAAATGGAGATA 1340	
RESULT 10			
US-10-649-273-21			
Sequence 21, Application US/10649273			
Publication No. US20040043407A1			
GENERAL INFORMATION:			
APPLICANT: Bristol-Myers Squibb Company			
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1			
FILE REFERENCE: D0073 CNT			
CURRENT APPLICATION NUMBER: US/10/649,273			
CURRENT FILING DATE: 2003-08-27			
PRIOR APPLICATION NUMBER: US 60/266,518			
PRIOR FILING DATE: 2001-02-05			
PRIOR APPLICATION NUMBER: US 10/067,443			
PRIOR FILING DATE: 2002-02-05			
PRIOR APPLICATION NUMBER: US 60/282,814			
PRIOR FILING DATE: 2001-04-10			
NUMBER OF SEQ ID NOS: 71			
SOFTWARE: PatentIn version 3.2			
SEQ ID NO 21			
LENGTH: 1387			
TYPE: DNA			
ORGANISM: homo sapiens			
US-10-649-273-21			
Alignment Scores:			
Pred. No.:			
Score:	6.62e-146	Length:	1387
	1213.50	Matches:	238
Percent Similarity:	90.15%	Conservative:	0
Best Local Similarity:	90.15%	Mismatches:	1

PRIOR APPLICATION NUMBER: US 10/067,443  
PRIOR FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 21  
LENGTH: 1387  
TYPE: DNA  
ORGANISM: homo sapiens  
US-10-651-722-21

## Alignment Scores:

Pred. No.:	6,82e-148	Length:	1387
Score:	1213.50	Matches:	238
Percent Similarity:	90.15%	Conservative:	0
Best Local Similarity:	90.15%	Mismatches:	1
Query Match:	97.86%	Indels:	25
DB:	19	Gaps:	1

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-651-722-21 (1-1387)

QY 1 LeuEuAlaLeuValGInGIyValSerAspPheLeuLeuGIyLysSerLeuAspIle 20  
Db 549 CTGTTGCATTAGTTCAAGAGAGTTTCAGATTTCGCTTCTGGAAGCTTTGGACATA 608  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
Db 609 GCACCAAGTGCATGCTTGACAGAGTGCAGAAAGACTTTCTTAATATAACATCCAGAG 668  
QY 41 CysSerThrMetSerGIyGIyLysValAlaIleGluHisLeuAlaLysGInGIyValAsnArgPhe 60  
Db 669 TGCTCCACCATGAGTGTGGGAAAGCCATAGACATTTGGCCAAACAGAAATGATT 728  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGIy 80  
Db 729 CATTTTGACATCAAACTCCCTTGATCATGCTAAATTTGGATTCTTTACTGGA 788  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGIyIle----- 98  
Db 789 CTTCAACACGTTACTGATTAATATATATGAAAAAGAAAGAGATATATTCTTA 848  
QY 98 ----- 98  
Db 849 ATTAGTAAGTTGAACAGATTAATATCTCGATTTGCTCTAAATTAAGCTGCTATTTC 908  
QY 99 -----GluLysGIyGInIleLeuSerSerAlaAlaAspIleAlaIleThrValGIn 115  
Db 909 TGCAGGTATGAGAGGGGCCAAATCCTGTCTTCAGCAGCAGACATTGCTCCACAGTACAG 968  
QY 116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGIn 135  
Db 969 CACACAAATGGCATGCTCTTGTGAAAAAGAACACATCGGGCTATTCTGTTTGTAAAGAG 1028  
QY 136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGIyGIyValAlaSerAsn 155  
Db 1029 AGAAGCTTTGTAACCTCAAAATATATGACATGCTGTCATGTCGTGTGCGAAGTAAAC 1088  
QY 156 PheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGInCysThrLeuLeuCys 175  
Db 1089 TTCTATATCCGAGAGCTCTGGAATTTTAAACAAACGACACAGAGCTTTGTGTGT 1148  
QY 176 ProProProArgLeuCysThrAspAsnGIyIleMetIleAlaIleTrpAsnGIyIleGluArg 195  
Db 1149 CCTCTCCACAGCTATGACATGATATATGATTCATTCATGATGATGATTTGAAAGA 1208  
QY 196 LeuArgAlaGIyLeuGIyIleLeuHisAspIleGluGIyIleArgTyrGluProLysCys 215  
Db 1209 CTACGTGTGTGCTTGGCATTTTATCATGACATAGAGGATCCGCTATGAAACCAAAATGT 1268  
QY 216 ProLeuGIyValAspIleSerTyrGIyValGIyValAlaSerIleLysValProGlnLeu 235  
Db 1269 CCTCTGAGTAGATATCAAAAGAGTGGAGAGCTTCCATTAAGTACCAACATTTA 1328

QY 236 LysMetGluIle 239  
Db 1329 AAAATGAGAGATA 1340

## RESULT 12

US-10-887-553A-1047  
Sequence 1047, Application US/10887553A  
Publication No. US20050085436A1  
GENERAL INFORMATION:

APPLICANT: Garza, Dan  
APPLICANT: Li, Hao  
TITLE OF INVENTION: Method to treat conditions associated  
with insulin signalling dysregulation  
FILE REFERENCE: 4-33262  
CURRENT APPLICATION NUMBER: US/10/887,553A  
CURRENT FILING DATE: 2004-07-08  
PRIOR APPLICATION NUMBER: 60/485,883  
PRIOR FILING DATE: 2003-08-07  
NUMBER OF SEQ ID NOS: 1208  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1047  
LENGTH: 1387  
TYPE: DNA  
ORGANISM: human  
US-10-887-553A-1047

## Alignment Scores:

Pred. No.:	6,82e-148	Length:	1387
Score:	1213.50	Matches:	238
Percent Similarity:	90.15%	Conservative:	0
Best Local Similarity:	90.15%	Mismatches:	1
Query Match:	97.86%	Indels:	25
DB:	22	Gaps:	1

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-887-553A-1047 (1-1387)

QY 1 LeuEuAlaLeuValGInGIyValSerAspPheLeuLeuGIyLysSerLeuAspIle 20  
Db 549 CTGTTGCATTAGTTCAAGAGAGTTTCAGATTTCGCTTCTGGAAGCTTTGGACATA 608  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
Db 609 GCACCAAGTGCATGCTTGACAGAGTGCAGAAAGACTTTCTTAATATAACATCCAGAG 668  
QY 41 CysSerThrMetSerGIyGIyLysValAlaIleGluHisLeuAlaLysGInGIyValAsnArgPhe 60  
Db 669 TGCTCCACCATGAGTGTGGGAAAGCCATAGACATTTGGCCAAACAGAAATGATT 728  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGIy 80  
Db 729 CATTTTGACATCAAACTCCCTTGATCATGCTAAATTTGGATTCTTTACTGGA 788  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGIyIle----- 98  
Db 789 CTTCAACACGTTACTGATTAATATATATGAAAAAGAAAGAGATATATTCTTA 848  
QY 98 ----- 98  
Db 849 ATTAGTAAGTTGAACAGATTAATATCTCGATTTGCTCTAAATTAAGCTGCTATTTC 908  
QY 99 -----GluLysGIyGInIleLeuSerSerAlaAlaAspIleAlaIleThrValGIn 115  
Db 909 TGCAGGTATGAGAGGGGCCAAATCCTGTCTTCAGCAGCAGACATTGCTCCACAGTACAG 968  
QY 116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGIn 135  
Db 969 CACACAAATGGCATGCTCTTGTGAAAAAGAACACATCGGGCTATTCTGTTTGTAAAGAG 1028  
QY 136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGIyGIyValAlaSerAsn 155  
Db 1029 AGAAGCTTTGTAACCTCAAAATATATGACATGCTGTCATGTCGTGTGCGAAGTAAAC 1088

```
Qy 156 PheTyrIleArgArgAlaLeuGluIleuThraenaIaThrGlnCysThrIleuLeuCys 175
Db 1089 TTTCTATATCCGACAGCTCTGAAATTTTAAACAAACGACACAGTGCATTGTGTGT 1148
Qy 176 ProProAlaArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArg 195
Db 1149 CTTCTCCACAGACTATGACATGATATGATTCATGATGGAATGGTATTGAAGA 1208
Qy 196 LeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProLeuCys 215
Db 1209 CTACGTGCTGCTGGCATTTTACATGACATAGAACGATCCGCTATGAACCAAAATGT 1268
Qy 216 ProLeuGlyValaAspIleSerIleGluValaGlyGluAlaSerIleValaProGluLeu 235
Db 1269 CCTCTTGAGATGACATATCAAAAGAGAGTGGAGAGCTTCATTAAGTACCAAAATTA 1328
Qy 236 LysMetGluIle 239
Db 1329 AAAATGGAGATA 1340

RESULT 13
US-10-012-140-6
; Sequence 6, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; FILE REFERENCE: 381552004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-012-140-6

Alignment Scores:
Pred. No.: 1,38e-146 Length: 1245
Score: 1203.00 Matches: 232
Percent Similarity: 98.33% Conservative: 3
Best Local Similarity: 97.07% Mismatches: 4
Query Match: 97.02% Indels: 0
DB: 15 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-012-140-6 (1-1245)
Qy 1 LeuLeuAlaLeuValaGlnGlyValaSerAspPheLeuLeuLeuGlyLysSerIleuAspIle 20
Db 526 CTGTGGCATATGATTTCAAGAGGATTTTCAGATTTTCGCTTTCGGAAGATCTTTCGACATA 585
Qy 21 AlaProGlyAspMetLeuAspIleValaAlaArgArgLeuSerIleuIleValaHisProGlu 40
Db 586 GCACCAAGTACACATGTCAGACAGAGTGCAGAGACATTTCTTAAATTAACATCCAGAG 645
Qy 41 CysSerThrMetSerGlyValaValaAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
Db 646 TGCCTCCACCATGAGTGGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATTT 705
Qy HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
Db 706 CATTTGACATCAAACTCCCTTCATCATCTAAATAATTTGATTTTCTTTACTGGA 765
```

```
Qy 81 LeuGlnHisValaThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100
Db 766 CTTCAACACGTTACTGATTAATAATATGAAGAAACAGGAAACAGAGAGAGATTTGAGAAG 825
Qy 101 GlyGlnIleLeuSerSerIleAlaAspIleAlaAlaIleValaGlnHisThrMetAlaCys 120
Db 826 GGGCAAAATCTGTCTTCACACACACATTCGTCACAGATACAGACACAAATGGCATGT 885
Qy 121 HisLeuValaLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
Db 886 CATCTTGCAAAAGAACACATCGGGCTATTCTTTTGTAGCAGAGACATTTGTTACT 945
Qy 141 GlnAsnAsnAlaValaLeuValaAlaSerGlyValaAlaSerAsnPheTyrIleArgArg 160
Db 946 CAAATTAATGACAGTACTGGTTCATCTGATGTGTGTGCAAGTAACTTATATCCGACAG 1005
Qy 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGlu 180
Db 1006 GCTCTGAAATTTTAAACAAACGACACAGTGCACATTTGTGTCTCTCCACACATA 1065
Qy 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 1066 TGCACCTGATATGCACTTATGATTCATGATGAAATGGTATTGAAGACTACGTGCTGCTTG 1125
Qy 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValaAsp 220
Db 1126 GGCATTTTACATGACATAGAGAGCATCCGCTATGAACCAAAATGTCTCTTGGAGTAGAC 1185
Qy 221 IleSerLysGluValaGlyGluAlaSerIleValaProGluLeuLysMetGluIle 239
Db 1186 ATATCAAAAGAGTGGAGAGAGCTTCATTAAGTACCAATTAATAATGAGAGATA 1242

RESULT 14
US-10-012-140-4
; Sequence 4, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; FILE REFERENCE: 381552004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1820
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (146)...(1390)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1820)
; OTHER INFORMATION: n = A,T,C or G
US-10-012-140-4

Alignment Scores:
Pred. No.: 2,41e-146 Length: 1820
Score: 1203.00 Matches: 232
Percent Similarity: 98.33% Conservative: 3
Best Local Similarity: 97.07% Mismatches: 4
```



Query Match: 97.02% Index: 0  
DB: 15 Gaps: 0  
US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-012-140-4 (1-1820)

```
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyValSerLeuAsp11e 20
DB 671 CTGTTGGCATTAATTCAGAGAGTTTCAGATTTTCGTTCTTGGAAAGTCTTTGACATA 730
QY 21 AlaProGlyAspMetLeuAspIlyValAlaArgArgLeuSerLeuIleValSHisProGlu 40
DB 731 GACCCAGGTACATGCTTGACAGAGTGCAGAGAACCTTCTTTAAATAACATCCAGAG 790
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuValAlaGlnGlyValAsnArgPhe 60
DB 791 TGCCTCCACCATGAGTGTGGGAAAGCCATAGACATTTGGCCAAACAGAAATATGATTT 850
QY 61 HisPheAspIleValProProLeuHisHisAlaIlyAsnGlyAspPheSerPheThrGly 80
DB 851 CATTTTGACATCAAACTCCCTTCATCATGCTAAATGATGATTTTCTTTTACTGGA 910
QY 81 LeuGlnHisValThrAspIlyValIleMetLysGlyGlyGluGluGlyIleGluLys 100
DB 911 CTTCAACACGTTACTGATTAATAATATGAAAAAGAGAAAGAGATTTGAGAG 970
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120
DB 971 GGGCAAACTCTGCTTCAGACAGACATTCCTCCACAGTACAGACACAAATGGCAGT 1030
QY 121 HisLeuValIlyAspIlyValIleLeuPheCysLysGlnArgAspLeuPro 140
DB 1031 CATCTTGAAAAAGAACACATCGGCTATCTGTTTGTAGACAGAGACTTGTACT 1090
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaAspAsnPheTyrIleArgArg 160
DB 1091 CAATAATATCAGATCGTGTGATGCTGTGCTGAGTACCTTATATCCGACGA 1150
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB 1151 GCTCTGAAATTTTAAACAAACGCAACACAGTGCATTTGTGTCTCTCCACGACTA 1210
QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1211 TGCATCATATATGCAATTAATGATGCAATGATGTAATAAGACTAGCGTGGCTTG 1270
QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyGluProLysCysProLeuGlyValAsp 220
DB 1271 GGCATTTTACATGACATAGAGGATCGCTATGAAACCAAAATGCTCTTGGAGTAGAC 1330
QY 221 IleSerLysGlyValIleGlyValaSerIleLysValProGlnLeuLysMetGluIle 239
DB 1331 ATATCAAAAGAGTTGGAGAGCTTCAATAAAGTACCAAAATTAATAATGAGATA 1387

RESULT 15
; Sequence 400. Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KETICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
```

```
; APPLICANT: OTSUKA, MOTOKYU
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 08435/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 400
; LENGTH: 2208
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-400
```

Alignment Scores:  
Pred. No.: 2,3e-127 Length: 2208  
Score: 1059.00 Matches: 211  
Percent Similarity: 88.28% Conservative: 0  
Best Local Similarity: 88.28% Mismatches: 4  
Query Match: 85.40% Indels: 24  
DB: 18 Gaps: 1

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-094-749-400 (1-2208)

```
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyValSerLeuAsp11e 20
DB 869 CTGTTGGCATTAATTCAGAGAGTTTCAGATTTTCGTTCTTGGAAAGTCTTTGACATA 928
QY 21 AlaProGlyAspMetLeuAspIlyValAlaArgArgLeuSerLeuIleValSHisProGlu 40
DB 929 GACCCAGGTACATGCTTGACAGAGTGCAGAGAACCTTCTTTAAATAACATCCAGAG 988
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuValAlaGlnGlyValAsnArgPhe 60
DB 989 TGCCTCCACCATGAGTGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATATGATTT 1048
QY 61 HisPheAspIleValProProLeuHisHisAlaIlyAsnGlyAspPheSerPheThrGly 80
DB 1049 CATTTTGACATCAAACTCCCTTCATCATGCTAAATGATGATTTTCTTTTACTGGA 1108
QY 81 LeuGlnHisValThrAspIlyValIleMetLysGlyGlyGluGluGlyIleGluLys 100
DB 1109 CTTCAACACGTTACTGATTAATAATATGAAAAAGAGAAAGAGATTTGAGAG 1168
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120
DB 1169 GGGCAAACTCTGCTTCAGACAGACATTCCTCCACAGTACAGACACAAATGGCAGT 1228
QY 121 HisLeuValIlyAspIlyValIleLeuPheCysLysGlnArgAspLeuPro 140
DB 1229 CATCTTGAAAAAGAACACATCGGCTATCTGTTTGTAGACAGAGACTGTTACT 1288
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaAspAsnPheTyrIleArgArg 160
DB 1289 CAATAATATGACATGCTGTGCATCTGTGTGTCGACAGTACTTCTGTATCCGACGA 1348
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB 1349 GCTCTGAAATTTTAAACAAACGCAACACAGTGCATTTGTGTCTCTCCACGACTA 1408
QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1409 TGCATCATATATGCAATTAATGATTTGCA----- 1435
QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyGluProLysCysProLeuGlyValAsp 220
DB 1436 -----TGATGCTCTTGGAGTAGAC 1456
```



OY 221 IleserLySGluValGlyGluAlaSerIleLyValProGlnLeuLyMetGluIle 239  
|||||  
Db 1457 AATCAAAAAGAGCTTGGAGAGCTTCATAAAAATGACCACAATTAATAAATGGAGATA 1513  
|||||

Search completed: November 11, 2005, 03:16:43  
Job time : 803.636 secs